Network-based modeling for risk assessment of infectious disease transmission

by

Md Mahbubul Huq Riad

B.S., Bangladesh University of Engineering and Technology, 2013

M.S., Kansas State University, 2017

AN ABSTRACT OF A DISSERTATION

submitted in partial fulfillment of the
requirements for the degree

DOCTOR OF PHILOSOPHY

Mike Wiegers Department of Electrical and Computer Engineering
Carl R. Ice College of Engineering

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2020

Abstract

Infectious disease modeling is crucial to optimize surveillance, preventative measures, and resource allocation. Simulation with infectious disease models is very convenient when the resource requirement for data collection and experimental studies are prohibitively high or even unethical. A vast number of approaches have been proposed to model infectious disease transmission from different perspectives. In this dissertation, we investigate network-based disease models for efficient resource allocation, effective mitigation measures, and accurate risk assessment. We also investigate a filtering-based parameter estimation and forecasting framework, usable when proper incidence data is available.

First, we provide a guideline for developing a network-based model and simulation framework for any infectious diseases. As an example, we provide a step-by-step method for developing a spatially explicit model for infectious diseases with host demographic data. We show how to devise effective mitigation strategies from simulation results using the spatially explicit model.

Our second contribution is developing a parameter estimation framework using a sequential Monte Carlo filter, a compartmental disease model, and historical incidence data. Parameter estimation for any infectious disease model is crucial for accurately informing resource allocation and control measures. Our method is particularly important for its adaptability to the availability of new incidence data of any epidemic. This parameter estimation framework is not limited to epidemic models; rather, it can be used for any systems with a state-space model.

Third, we propose an ensemble Kalman filter that provides dual state-parameter estimates for infectious diseases. As an online inferential method, the ensemble Kalman Filter can perform real-time forecast during an outbreak. The framework is capable of accurate short to mid-term forecasts.

Fourth, we develop a risk assessment framework for infectious diseases with a comprehensive two-layer network— a permanent layer representing permanent contacts among individuals, and a data-driven layer for temporary contacts due to movements. We combine the two-layer network with a compartmental model and implement a Gillespie algorithm to identify the disease evolution and assess the spatial spreading risk. The proposed risk assessment framework suggests some focal points (spatial) for disease preparedness, providing critical directions to inform interventions in the field.

Finally, we investigate the strong correlation of the arthropod abundance and host interaction with vector-borne pathogen transmission, and we developed a risk assessment framework using climate (average temperature and rainfall) and host demographic (host density and movement) data, particularly suitable for regions with unreported or under-reported incidence data. This framework consisted of a spatiotemporal network-based approach coupled with a compartmental disease model and a non-homogeneous Gillespie algorithm. We have identified the spatiotemporal suitability map, the spatial risk map, the significant-incidence window, and peak incidence period. The outcomes of the framework comprise of weather-dependent spatiotemporal suitability maps and probabilistic risk maps for spatial infection transmission. This framework is capable of vector-borne disease risk assessment without historical incidence data and can be a useful tool for preparedness with accurate human movement data.

Network-based modeling for risk assessment of infectious disease transmission

by

Md Mahbubul Huq Riad

B.S., Bangladesh University of Engineering and Technology, 2013

M.S., Kansas State University, 2017

A DISSERTATION

submitted in partial fulfillment of the
requirements for the degree

DOCTOR OF PHILOSOPHY

Mike Wiegers Department of Electrical and Computer Engineering
Carl R. Ice College of Engineering

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2020

Approved by:

Major Professor
Caterina Maria Scoglio

Copyright

Abstract

Infectious disease modeling is crucial to optimize surveillance, preventative measures, and resource allocation. Simulation with infectious disease models is very convenient when the resource requirement for data collection and experimental studies are prohibitively high or even unethical. A vast number of approaches have been proposed to model infectious disease transmission from different perspectives. In this dissertation, we investigate network-based disease models for efficient resource allocation, effective mitigation measures, and accurate risk assessment. We also investigate a filtering-based parameter estimation and forecasting framework, usable when proper incidence data is available.

First, we provide a guideline for developing a network-based model and simulation framework for any infectious diseases. As an example, we provide a step-by-step method for developing a spatially explicit model for infectious diseases with host demographic data. We show how to devise effective mitigation strategies from simulation results using the spatially explicit model.

Our second contribution is developing a parameter estimation framework using a sequential Monte Carlo filter, a compartmental disease model, and historical incidence data. Parameter estimation for any infectious disease model is crucial for accurately informing resource allocation and control measures. Our method is particularly important for its adaptability to the availability of new incidence data of any epidemic. This parameter estimation framework is not limited to epidemic models; rather, it can be used for any systems with a state-space model.

Third, we propose an ensemble Kalman filter that provides dual state-parameter estimates for infectious diseases. As an online inferential method, the ensemble Kalman Filter can perform real-time forecast during an outbreak. The framework is capable of accurate short to mid-term forecasts.

Fourth, we develop a risk assessment framework for infectious diseases with a comprehensive two-layer network— a permanent layer representing permanent contacts among individuals, and a data-driven layer for temporary contacts due to movements. We combine the two-layer network with a compartmental model and implement a Gillespie algorithm to identify the disease evolution and assess the spatial spreading risk. The proposed risk assessment framework suggests some focal points (spatial) for disease preparedness, providing critical directions to inform interventions in the field.

Finally, we investigate the strong correlation of the arthropod abundance and host interaction with vector-borne pathogen transmission, and we developed a risk assessment framework using climate (average temperature and rainfall) and host demographic (host density and movement) data, particularly suitable for regions with unreported or under-reported incidence data. This framework consisted of a spatiotemporal network-based approach coupled with a compartmental disease model and a non-homogeneous Gillespie algorithm. We have identified the spatiotemporal suitability map, the spatial risk map, the significant-incidence window, and peak incidence period. The outcomes of the framework comprise of weather-dependent spatiotemporal suitability maps and probabilistic risk maps for spatial infection transmission. This framework is capable of vector-borne disease risk assessment without historical incidence data and can be a useful tool for preparedness with accurate human movement data.

Table of Contents

List of Figures

List of Tables

Acknowledgments

I want to thank my advisor Dr. Caterina Scoglio and Dr. Lee Cohnstaedt for their continuous support, guidance, and enormous endeavor in my academic progression. They always allowed me to share my thoughts and problems — academic or otherwise– and tried their best to guide me in overcoming them. It would be challenging for me to finish this dissertation without their technical and editorial advice.

I owe my profound gratitude to my committee members, Dr. Don Gruenbacher and Dr. Hongyu Wu, for their time and support. Sincere gratitude to Dr. David Renter for serving as the chairperson of the examining committee for this doctoral degree.

Heartfelt thanks to my loving wife, Shiba, for her endless love and support. Her unconditional and unending support kept me going with my work and studies. I express my deepest gratitude to my beloved sister and niece, Disha, for their unending support and encouragement all through my career. Words are not enough to thank my beloved parents for their eternal love and endless prayers. I take this opportunity to thank my family members for moral support and devoted pieces of advice. Last but not least, all thanks to the Almighty for allowing me to pursue my dream.

Dedication

To my parents

Preface

This dissertation, with the title "Network-based modeling for risk assessment of infectious disease transmission" is submitted for the degree of Doctor of Philosophy in the Department of Electrical and Computer Engineering at Kansas State University. The research has been performed under the supervision of Prof. Caterina Scoglio. Most of the work is comprised of the following set of published or submitted peer-reviewed journal and conference papers:

Peer-reviewed journal papers

- Mahbubul H Riad, Lee W Cohnstaedt, and Caterina M Scoglio. Risk assessment of vector-borne disease transmission using spatiotemporal network model and climate data with an application of dengue in Bangladesh. Submitted for publication.

- Mahbubul H Riad, Musa Sekamatte, Felix Ocom, Issa Makumbi, and Caterina M Scoglio. Risk assessment of Ebola virus disease spreading in Uganda using a two-layer temporal network. *Scientific reports*, 9(1):1–17, 2019. (Impact factor: 4.120)

- Mahbubul H Riad, Caterina M Scoglio, D Scott McVey, and Lee W Cohnstaedt. An individual-level network model for a hypothetical outbreak of Japanese encephalitis in the USA. *Stochastic environmental research and risk assessment*, 31(2):353–367, 2017. (Impact factor: 3.09)

- Musa Sekamatte, Mahbubul H Riad, Tesfaalem Tekleghiorghis, Kenneth J Linthicum, Seth C Britch, Juergen A Richt, JP Gonzalez, and Caterina M Scoglio. Individual-based network model for Rift Valley fever in Kabale district, Uganda. *PloS one*, 14(3): e0202721, 2019. (Impact factor: 2.870)

- Caterina M Scoglio, Claudio Bosca, Mahbubul H Riad, Faryad D Sahneh, Seth C Britch, Lee W Cohnstaedt, and Kenneth J Linthicum. Biologically informed individual-based network model for Rift Valley fever in the US and evaluation of mitigation strategies. *PloS one*, 11(9):e0162759, 2016. (Impact factor: 2.870)

Peer-reviewed conference papers

- Mahbubul H Riad, Caterina M Scoglio, D Scott McVey, and Lee W Cohnstaedt. Estimation of parameters and basic reproductive ratio for Japanese encephalitis transmission in the Philippines using a sequential Monte Carlo filter. In *2017 IEEE Conference on Control Technology and Applications (CCTA)*, pages 668-673. IEEE, 2017 (Acceptance rate: 63%)

- Mahbubul H Riad, Caterina M Scoglio, Lee W Cohnstaedt, and D Scott McVey. Short-term forecast and dual state-parameter estimation for Japanese encephalitis transmission using ensemble Kalman filter. In *2019 American Control Conference (ACC)*, pages 3444–3449. IEEE, 2019. (Acceptance rate: 64%)

Chapter 1

Introduction

1.1 Introduction

Infectious disease models provide guidelines in efforts focusing on understanding, risk assessment, and control of current and potential outbreaks. A crucial factor in the infectious disease model is incorporating the impact of pathogen and host-level interactions and ecological, social, and demographic factors[1–3]. Mathematical modeling and simulation are very convenient when the resource requirement for data collection and experimental studies are prohibitively high. A vast number of approaches have been proposed to model infectious disease transmission from different perspectives. Some of these models are mechanistic models of disease dynamics (compartmental models, differential equation models, complex network models)[4–10], statistical methods (regression model, hidden Markov model)[11–16], and data-driven methods for the forecasting (machine learning model)[17–20]. These models provide important guidelines about disease transmission, spatial patterns, and mitigation measures.

While the history of the infectious disease model dates back to the eighteen century, most of them focused on the statistical and mechanistic models[3]. While statistical methods have the potential for rapid assessment of emerging situations, their success is crucially dependent on the correctness of the data and the construction of statistical models to effectively capture disease characteristics[3;21]. A limitation of the early statistical methods was the inability to

record explicitly the disease characteristics rather than observing some possible indicators of the disease. Mechanistic models, on the other hand, provides crucial insights about critical factors in disease transmission systems[22]. Compartmental models can be explored using powerful analysis techniques for ordinary or partial differential equations. Under the mean-field approximation, assuming that the population is perfectly mixed and that every susceptible has the same probability of becoming infected, the probabilities are equated to the expected (mean) values of the corresponding variables in the population.

However, due to the complexity and the stochasticity of the disease transmission, most available compartmental models are often only qualitative[3;23]. The qualitative results can be very useful, especially when a large portion of the population is infected. However, they cannot capture the comprehensive epidemiological realism. The strict mean-field approximation and homogeneous mixing/contacts of the population in the differential equation models are relaxed in complex network mechanistic models, which are alternatively called as network-based models[3;10]. Contacts between individuals are influenced by a broad spectrum of factors characterizing the transmission mechanisms of a particular disease. Understanding the role of intricate contact patterns is of utmost importance to public-health measures and policies for controlling disease outbreaks. Vaccination, quarantine, and use of antiviral drugs on targeted parts of the population have to be carefully designed to efficiently combat an emerged epidemic. Poor understanding of the infectious disease dynamics as these emerge due to heterogeneous contact interactions may result in serious negative consequences. Over the last years, there has been an intense effort in studying the interplay between the emergent dynamics of infectious diseases and the underlying topology of the transmission network.

Therefore, network-based models have become very popular in recent times due to their capability to incorporate heterogeneous contact structure of individual nodes. Extensive research during the past decade has been devoted to capturing the network's role and the interaction of different system components on infectious disease transmission dynamics[8;10;24–28]. However, there are various topics in network-based models that need further exploration. In this dissertation, we focus on 1) providing explicit guidelines on developing spatially explicit network-based models and suggesting mitigation measures, 2) exploring applications of well-

known parameter estimation and forecasting methods in the context of stochastic disease transmission model, and 3) assessing the risk of infectious disease transmission using realistic network-based models.

1.2 Motivation

A crucial factor for the network-based model is the identification of valuable, and relevant information in the model. However, this information is in general particular to the a disease, a geographic location, demographic information, and environmental factors. Incorporating irrelevant information will complicate the model without providing significant improvement in the existing body of knowledge. Therefore, in this work, we provide guidelines for formulating a network-based model with the application to specific diseases. Simulation results from the network-based model are used to reinforce the knowledge of the disease dynamics as well as suggesting effective mitigation measures. These network-based models will be useful for public health policymakers when realistic information about the network and disease transmission parameters are incorporated.

The realistic disease transmission parameters can be derived from the disease incidence data for a specific geographic location. Essential methods for deriving the parameter are based on filtering approaches such as the adaptation of particle filters and Kalman filters. To estimate realistic parameters, we have developed a parameter estimation framework using sequential Monte Carlo filter from disease incidence data and have demonstrated its effectiveness with an application. This parameter estimation framework can capture the temporal variation of the parameter for specific diseases and incidence data. Once these realistic parameter estimates are used in the network-based model, the simulation results will provide guidelines for public health people in effective resource allocation.

We have also developed a dual-state parameter estimation framework from historical incidence data adapting the ensemble Kalman filter for disease transmission models. The prediction/forecasting of infectious disease is highly important for preparedness. The dual-state parameter estimation framework can be used for simultaneously forecasting and estimating

model parameters. The framework is capable of accurate short and mid-term forecasts for infectious diseases.

Risk assessment of infectious disease plays a vital role in disease preparedness. The contact and the connectivity among the host population call for a risk assessment method using a network-based model. Therefore, we have developed network-based risk assessment models for infectious disease. For diseases that are directly transmitted from an individual to another, contact and movement structure plays a vital role in disease transmission. However, the contact network topology keeps changing due to the movement of individuals. To account for this changing network topology, we have proposed a two-layer temporal network for risk assessment framework. The network has a permanent layer representing permanent contacts among individuals and a data-driven temporal network for human movements. We propose a Gillespie algorithm with a compartmental model to simulate the evolution of disease spreading as well as to evaluate the risk throughout our network.

Finally, we have developed a risk assessment framework for vector-borne diseases, whose transmission is dependent on a vector population. Vector-borne disease risk assessment is crucial to optimize surveillance, preventative measures (vector control), and resource allocation (medical supplies). High arthropod abundance and host interaction strongly correlate to vector-borne pathogen transmission. Increasing host density and movement increases the possibility of local and long-distance pathogen transmission. We developed a risk assessment framework using climate (average temperature and rainfall) and host demographic (host density and movement) data, particularly suitable for regions with unreported or under-reported incidence data.

1.3 Broader impacts

This dissertation provides a generalized method for developing a network-based models by incorporating relevant components for a spatially explicit network. Simulation results demonstrate the usefulness of the model for effective resource allocation and suggesting mitigation measures.

We investigate a method to derive realistic parameter values to be used in the network-based model. The realistic parameters provide confidence for simulation results to be used for practical purposes such as resource allocation and mitigation measures suggestion. We also develop a short-term and mid-term forecasting framework, which is capable of providing a picture of the severity of an outbreak in the near future. The applications of the parameter estimation and forecasting framework are not limited epidemic model and can be used to other models in engineering, biology, and data science.

We develop risk assessment frameworks for infectious diseases in the absence of incidence data. We develop the risk assessment framework for both host-host and host-vector-host transmission. These frameworks are based on the host demographic data (density, movements) and climate data, and are a useful tool for providing guidelines to public health personnel for new infectious diseases and in locations with unreported and under-reported incidence data.

1.4 Contributions

Our contributions can be summarized as follows.

- Demonstrating the method for developing a contact network and simulating the infection spreading within a network with an application to Rift valley fever and Japanese encephalitis transmission.
- Suggesting mitigation interventions during an outbreak from network-based model's simulation results.
- Developing a parameter estimation framework using sequential Monte Carlo filter and using historical incidence data.
- Estimating basic reproductive ratio for infectious disease from historical incidence data.
- Adapting the ensemble Kalman filter (EnKF) for dual state-parameter estimation.

- Performing short- and mid-term retrospective forecasting for Taiwan Japanese encephalitis incidences and evaluate the accuracy of the forecast from the developed framework.
- Proposing a two-layer temporal network with a static and a temporal/dynamic layer incorporating both permanent and temporal contacts among individuals.
- Combining the Gillespie algorithm with the two-layer network to develop a framework for risk assessment and applying the framework for the Ebola risk assessment in Uganda.
- Modeling vectorial capacity for vector-borne disease transmission using climate data.
- Developing a spatiotemporal network model and a risk assessment framework for vector-borne diseases implementing a non-homogeneous Gillespie algorithm.
- Creating spatiotemporal suitability maps and spatial risk maps for dengue spreading in Bangladesh.

1.5 Organization

The concepts of network-based model and the guidelines for spatially explicit network formulation and simulation framework are presented in Chapter 2 and Appendix A with applications for Rift valley fever and Japanese encephalitis, respectively. We describe the parameter estimation framework using sequential Monte Carlo filter and its application for parameter estimation and forecasting for Japanese encephalitis in Philippines in Chapter 3. Chapter 4 presents the dual state-parameter estimation and forecasting framework using the ensemble Kalman filter, with an application for Japanese encephalitis forecast in Taiwan. In chapter 5, we introduce a two-layer temporal network for risk assessment of infectious diseases directly transmitted via physical contact. The application of the framework is demonstrated on a possible Ebola outbreak in Uganda from the neighbouring Democratic Republic of Congo. Finally, a risk assessment framework of vector-borne diseases is presented in Chapter 6 with

host demographic information and climate data. Closing remarks with future direction on this research are reported in Chapter 7.

1.6 Background

1.6.1 Infectious disease transmission over complex networks

In this section, we explain the infectious disease transmission over a complex network. As an example, we have used *Susceptible-Exposed-Infected-Recovered (SEIR)* compartmental model to describe the basics of infectious disease spreading in a network.

The network of a selected population with N individuals can be represented by $G = \{V, E\}$, where V represents the set of nodes, and E is the set of edges—possible means of infection transmission between the nodes. This network is called the contact network as it expresses the possible contact patterns of the nodes. The contact network is mathematically expressed as the adjacency matrix $A = [a_{kl}] \in R^{NXN}$ is defined with the elements $a_{kl} = 1$ if and only if $(k, l) \in E$ else $a_{kl} = 0$.

Nodes influence each other through statistically independent pairwise interactions in most network-based models. The combined state of all nodes in a network can be described as a random variable $X_N(t) = [x_1(t), x_2(t), ..., x_i(t)]$, where $x_i(t)$ is the state (compartment) of node i at time t. The transition time from one state to another is expressed as an exponential distribution with a transition rate $\sigma_n(x_n \rightarrow J)$, where J is the destination state after the transition. This transitions can be node-based (dependent only on the node state $x_i(t)$) or edge-based (dependent on the combined network state $X_N(t)$). The transition from susceptible-exposed is edge-based, and a susceptible node becomes infected through interaction with infected neighbors in the network. The transition from susceptible to infected and infected to recovered happen automatically after a certain time. These inter-state transition times are a random variable that can have any distribution. However, in the Markovian process, these transition times are exponentially distributed with specific rates. Sahneh et al. developed the *generalized epidemic modeling framework*(GEMF) for stochastic spread-

ing processes over complex networks based on these independent pairwise interactions for Markovian processes[29;30]. The overall structure of a network-based model is represented in Figure 1.1.

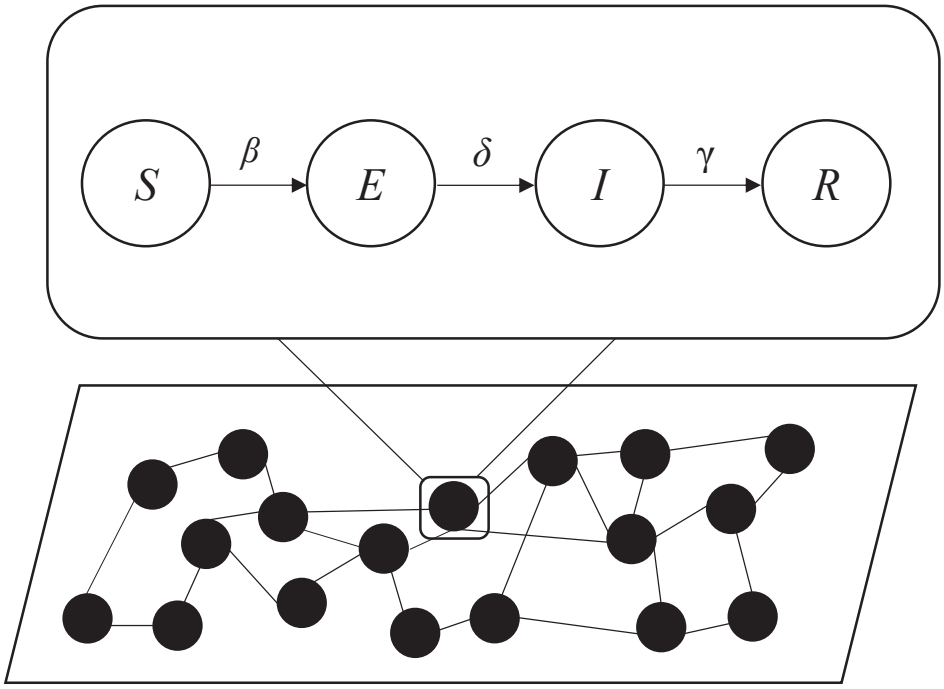


Figure 1.1: *Diagram of a network model that consisting of a SEIR model and a network. The bottom part within the rectangle represents the network, where each black dot is a node. The line between two nodes represents the link— possibility of pathogen transmission from an infected to a susceptible node. Circles in the top part represent the four compartments susceptible (S), exposed (E), infectious (I), and recovered (R) of a node, and arrows between the compartments show the direction of transition for each node with rates driven by parameters β (infection rate), δ (incubation rate), and γ (recovery rate).*

In *SEIR* compartmental model, each network node can assume one of the four compartments/states. The state occupancy probability after a time Δt can be expressed as equations (1.1- 1.3).

$$Pr[x_i(t+\Delta t)=2|x_i(t)=1,X_N(t)]=\beta Y_i\Delta t+o(\Delta t) \tag{1.1}$$

$$Pr[x_i(t+\Delta t)=3|x_i(t)=2,X_N(t)]=\delta\Delta t+o(\Delta t) \tag{1.2}$$

$$Pr[x_i(t+\Delta t)=4|x_i(t)=3, X_N(t)] = \gamma\Delta t + o(\Delta t) \tag{1.3}$$

In these equations, $x_i(t+\Delta t) = 1$, 2, 3, and 4 express the probability of node i occupying the susceptible, exposed, infected or removed state at time $(t+\Delta t)$, respectively. $X_N(t)$ is the combined network state at time t. The transition rate from susceptible to an exposed state is an edge-based transition. We express this parameter with an infection rate β. Y_i is the set of infected neighbors of node i within the network at time t. The parameter δ is the incubation rate, which governs the transition from exposed to infected state. The transition from infected to removed state is expressed with the recovery rate γ. Incubation rate δ and recovery rate γ are node-based transition rates.

A simulation tool GEMFsim was developed for numerical simulation by implementing the Gillespie algorithm with the Markovian process described in equations 1.1-1.3[30]. The Gillespie algorithm samples the earliest event among a set of independent (i.e., inter-state transitions) events with exponentially distributed occurrence time[31]. For example, node n will make a transition from state i to state j at a random time T_n which is exponentially distributed with a rate r_n. Since the inter-state transition of a node does not affect the transition of other nodes, the transition time for each node can be generated by drawing a random value from its corresponding distribution. We get a sequence of events by arranging these transition times in increasing order. The Gillespie algorithm samples the time for the earliest event by going through all the ongoing processes and the node that makes the transition. The GEMFSim tool was developed to simulate the stochastic spreading process with the Gillespie algorithm, as it is capable of simulating the Markovian dynamics of a complex system. It provides a simulation tool for exact, continuous-time numerical simulation of the spreading process over a complex network. Therefore, GEMFSim can be used for understanding the spreading dynamics of infectious diseases within a complex network-based model.

1.6.2 Parameter estimation and forecasting with filtering methods

The primary purpose of the filtering methods in epidemic models is to characterize the epidemiological system in the future from the initial information. The states of an epidemiological system X_k, at time k is dependent on the observations up to time k. A variety of methods in estimation theory enable the recursive estimation of system state variables and the inference of model parameters[32]. Filtering methods, also known as sequential data assimilation, have been widely used in different engineering and epidemiological designs and forecasts. The ability of an epidemiological model to make accurate predictions depends on the extent to which the model represents real-world transmission dynamics as well as the proper specification of model parameters and initial conditions[33]. Filtering methods use the available data to recursively inform and train the model so that current conditions are better depicted and evolving outbreak characteristics (i.e., the trajectory of the epidemic curve) are better-matched[34]. Some examples of filtering methods are different kinds of particle and Kalman filters[32;35], which are extensively used for parameter estimation and forecasting.

Chapter 2

Individual-based network model for Rift Valley fever[1]

2.1 Introduction

Rift Valley fever (RVF) is a zoonotic mosquito-borne disease caused by the Rift Valley fever virus (RVFV; *Phlebovirus*: Bunyaviridae). It severely affects ungulate livestock and wildlife but can also affect humans in RVF-endemic regions of sub-Saharan Africa and parts of the Arabian Peninsula[8;37]. Major RVF outbreaks have been reported in Egypt (1977, 2003), Kenya (1997, 1998, 2006, 2007), Tanzania (2007), Somalia (2007), Saudi Arabia and Yemen (2000-2001), Sudan (2007), Senegal (2013-2014), Mauritania (2010, 2012, 2013- 2014), Uganda (2016), and Niger (2016)[38–48]. Potential economic impact and public and veterinary health burdens due to RVF outbreaks have been documented[38;47–50]. Persistent heavy rainfall causing flooding is the most prominent precursor of RVF epizootics in East Africa, due to flooded- ground pools stimulating massive emergence of transovarially RVFV-infected Aedes mosquitoes[51;52]. The transmission cycle of RVFV initiates as the virus is introduced into livestock by competent mosquitoes during blood feeding[29;37;53]. However, the West Africa epizootic regions do not experience transmissions linked to elevated rainfall[54]. In these areas,

[1] This chapter is a reformatted and slightly modified version of our published article[36]

RVFV is most likely spread via movements of infected livestock from endemic areas. Livestock trading across different market areas may include infected cows that could disperse the virus in the presence of competent mosquitoes[55]. Patterns of recent RVF activity in Uganda support the hypothesis of RVFV spread linked to the cattle trade[56]. This event in Uganda underscored the need to develop effective operational surveillance and mitigation strategies to reduce or prevent spread among cattle operation locations. Mathematical/epidemic models offer the possibility to investigate RVFV and other infectious disease dynamics through time, and may be used to devise mitigation strategies[7]. The potential impact of an RVFV outbreak can be quantitatively assessed from simulations using epidemic models. The importance of space in RVF endemicity in West Africa was demonstrated by placing a mosquito habitat under surveillance to find the triggering point for an RVF epidemic[57]. Models showed animals could infect humans and mosquitoes; however, humans cannot infect mosquitoes or livestock[58]. A Bayesian spatial model for RVF spreading was proposed to investigate environmental drivers that alter host and vector distributions[59]. In Kenya, an ecological niche model was formulated to predict the distribution of RVF vector species under climate change[60]. An individual-level network model was proposed to demonstrate the effect of network topologies based on inter-farm cattle movement in the United States[7]. Two separate kernel functions — exponential and power-law kernels— were used to model cattle movement within and among farms in Riley County, Kansas. Between simulations with two kernel functions, widespread epizootics from the power-law model were revealed, because cows were allowed to move to distant farms. In contrast, the exponential model greatly restricted cattle movement to more proximal farms, reducing the spread of the virus. In this study, we develop a network-based epidemic transmission model to perform simulations. Simulation results provide an opportunity to investigate patterns of RVFV across locations in the Kabale District, Uganda. We build upon a previous individual-based network model to investigate RVFV epidemiology in the Kabale District using 2012 livestock data from UBOS[57]. This model considers livestock as a spatially explicit factor in an individual-based network representing different locations with the specific mosquito and environmental factors. Our goal is to investigate changes in the epidemic size (total number of infected cows) for varying mosquito abundance, different

initial conditions (single- or multiple-outbreak locations), cattle breed (indigenous or exotic), and cattle movement. We are able to suggest several mitigation strategies to check/reduce RVFV spread using simulation results from the individual-based network model.

2.2 Materials and method

2.2.1 Modeling framework

An RVFV modeling framework consists of two parts, a node transition graph, and a contact network. The node transition graph consists of four compartments— susceptible (S), exposed (E), infectious (I), and recovered (R). Each individual cow can be in only one of these four compartments, and rates of transitions between compartments are driven by parameters β (infection rate), δ (incubation rate), and γ (recovery rate). Figure 1.1 represents the spread model's conceptual core, showing the sequence of the progression of the RVFV infection in a cow (node) through four compartments.

In the network model, the infection can spread if a susceptible node (i.e., a susceptible cow) is in physical proximity with at least one infectious node. Specifically, one infectious cow (node 1) will be able to transmit RVFV to a susceptible cow (node 2) only if there are enough RVFV-competent mosquitoes to first bite the infectious cow (node 1) then, after an appropriate period for the virus to disperse and replicate in the mosquito, bite a susceptible cow (node 2)[7]. As stated before, links between cows in the network represent the possibility of virus transfer via mosquitoes once cows are in physical proximity for a sufficient period.

We explicitly model cows, and mosquitoes are included in an aggregated way with a transmission parameter from an infected animal to a susceptible one. This transmission parameter is directly proportional to vectorial capacity, which includes mosquito abundance, survival rate, vector competence, and feeding patterns[7]. Once a susceptible (S) node is in the physical proximity of an infectious node, virus transfer takes place with infection rate β and move the cow into the exposed (E) compartment. If a susceptible cow has Y_i infectious neighbors, then the probability of the susceptible cow to receive a virus transmission is

βY_i. Therefore, the total rate at which susceptible cows become infected is proportional to the number of infectious cows in the neighborhood and the vectorial capacity of available mosquito vectors. The cow's transition from the exposed compartment (E) to the infectious (I) compartment takes place at rate δ. It represents the time the pathogen will take, once it entered into the host body, to replicate enough for the cow to become infectious – i.e., capable of infecting a naive mosquito. Infectious cows finally transferred to the recovered/removed compartment (R) with rate γ. We do not consider disease-induced mortality; the endpoint in the simulation for an individual cow (node) is reached when it entered the R compartment.

Parameters δ and γ are specified according to the literature. For our simulations, we invariably use the value of δ=0.33 day^{-1} (3-day incubation period) and γ =0.14 day^{-1} (7-day recovery period)[7]. Infection rate β depends upon vector abundance as well as various environmental factors and cannot be expressed with a single value. Therefore, we use a range of β to explore various magnitudes of environmental factors as well as mosquito abundance. The infection rate is proportional to the realized vectorial capacity of competent mosquito species likely to be present in the study area.

After developing the individual-based *SEIR* network model for the Kabale District, we carry out extensive simulations using GEMFsim developed by the Network Science and Engineering (NetSE) group at Kansas State University[29;30]. We use the GEMFsim tool for simulation because it is an individual-based model, which provides more accurate predictions than meta-population models[7;29;30].

2.2.2 Geographic structure and movement in the cattle contact network (CCN)

We model the cattle movement network based on the local trading system for the Kabale District while considering two different networks depending upon the relative susceptibility of exotic and indigenous cattle. The cattle contact network consists of 20,806 cows (N) unevenly distributed across 22 locations in the Kabale District of Uganda in 2012 (Table 2.1), which is approximately 1,679 km^2 (648 sq mi) in the western region of Uganda (UBOS).

Table 2.1: *Cows in different locations in the Kabale District; this data set was derived from the UBOS Statistical Report 2012, Kabale District.*

Location	Number of Exotic Cows	Number of Indigenous Cows	Total
Bubale	1721	1580	3301
Bufundi	74	804	878
Buhara	215	837	1052
Bukinda	61	268	329
Butanda	24	403	427
Hamurwa	267	1083	1350
Hamurwa T/C	116	582	698
Ikumba	141	845	986
Kabale Municipality	336	600	936
Kaharo	87	578	665
Kamuganguzi	367	526	893
Kamwezi	187	1623	1810
Kashambya	68	721	789
Katuna T/C	304	271	575
Kitumba	187	692	879
Kyanamira	361	719	1080
Maziba	141	427	568
Muhanga T/C	42	276	318
Muko	38	872	910
Rubaya	180	1008	1188
Ruhija	8	382	390
Rwamucucu	71	713	784
Total	4996	15810	20806

Locations are represented by data from a sub-county, municipality (Kabale), or town council (Hamurwa, Muhanga, and Katuna) boundary (Table 2.1). We extracted the longitude and latitude of the centroid of each location from Google Maps to display in a GIS, as shown in Figure 2.1. We have further addressed each of them only by location without any distinction.

To capture the actual movement of cows in Uganda, we treat contact among cows (not physically; instead, it was implicit contact via mosquitoes) differently depending on a geographic scale. Cows are assumed to move freely within each location, while their movement is restricted between locations. We assume each cow had equal connection probability to all other individual cows in that location via mosquitoes because of their proximity. We found

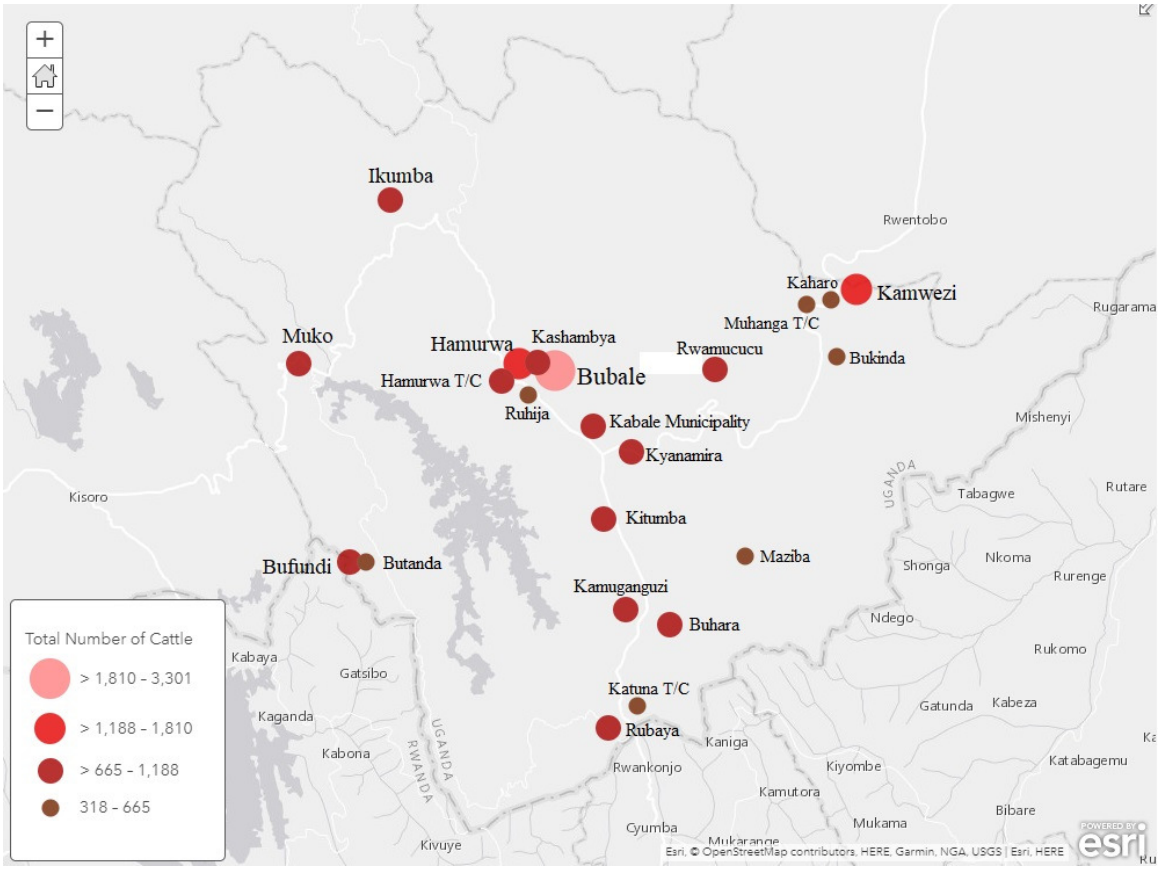


Figure 2.1: *Locations of cattle contact networks in the Kabale District; circles represent the center of each location. Circles are color-coded to the total number of cattle in each location. The bigger size of the circle represents higher numbers of cattle within a location.*

that an Erdos-Renyi network best represents this relationship among cows within locations, where each cow has an equal probability of connectivity (we assumed probability 0.7 for a connected network) to any other cow[30;61].

Transmission of RVFV from one location to another can happen via the movement of cows for economic reasons, most commonly through sales at local market places. Therefore, contact among cows, i.e., the possibility of virus transfer, is weighted in proportion to the distances between locations for the local trading system. We accomplish this weighting with an exponential distance kernel, expressed as e^{-kd}. k is a constant, which scales the probability of cows from different locations to be in contact and has a unit km^{-1}, and d is the distance between the origin and destination locations. We assume three different values of k, 0.001, 0.01, and 0.1 to reflect low, medium, and high movement probability. However,

the network is valid for any value of k. We model potential transmissions of RVFV that resulted from movement; therefore, an infected transferred cow to a new location can infect others via local mosquitoes at the destination location.

We visualize 20,806 cows across the 22 locations using the network visualization software Gephi[62] in Figure 2.2, but scale cattle population sizes across the network by a factor of $\frac{1}{20}$ for clarity. It is important to note that scaling is only used for visualization and not model simulations, which are performed with the full value of N.

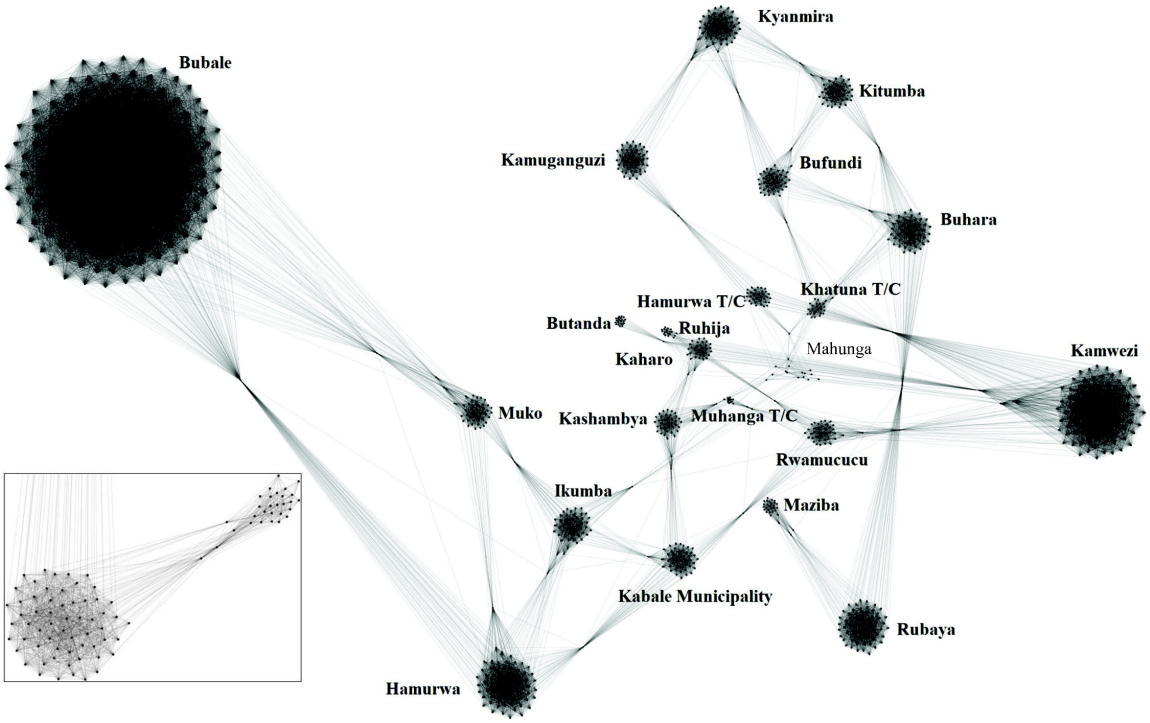


Figure 2.2: *Overall structure of the network; dense circular groupings of black dots represent different locations. The inset shows a close-up of two such groupings and one possible arrangement of links within and between them. Long black lines connect some locations, representing potential movement-related connections, and thus opportunities for mosquito-mediated transmission of RVFV between cows from different locations. The inset expands a small portion of the contact network showing the dense circular masses are made up of small black circles, each representing 20 cows and corresponds to the nodes shown in the representative contact network. Likewise, the black lines among these nodes represent possible connections within and between locations in the inset.*

2.2.3 Cattle contact network scenarios

Cases in the literature indicated exotic cows showed more susceptibility to RVF than indigenous cows. In Kenya, indigenous cows exhibited mild symptoms from RVFV infection, and these cows might develop lower viremia, which could significantly affect the transfer of the virus to mosquito vectors[63]. However, we do not have specific information on the relative susceptibility of indigenous compared to exotic cows for the Kabale District in Uganda. This relative susceptibility can vary with the breed as well as origin. Therefore, we assume two different network scenarios to capture the relative susceptibility of exotic versus indigenous cow breeds while performing simulations with GEMF: a homogeneous and a heterogeneous network. In the homogeneous network, we assume that all cows, indigenous or exotic, have the same susceptibility to RVFV. Therefore, we use the total number of cows in each location rather than differentiating them in two different categories.

In the heterogeneous network, we assume exotic cows are more susceptible to RVFV than indigenous cows. Lacking proper knowledge about the relative susceptibility, we assume if exotic cows have a susceptibility ζ, then indigenous cows had a susceptibility of $\mu\zeta$, where μ had a value between zero and one. $\mu = 1$ means a completely homogeneous network, while $\mu = 0$ means a network where indigenous cattle are immune to the RVFV pathogen. An increase in the value of μ from the minimum would increase network homogeneity and vice versa. For simulation purposes, we assume $\mu = 0.7$, which indicates thirty percent less susceptibility of the indigenous cows than exotic. However, we use this value to demonstrate the effects of heterogeneity in RVFV transmission in a qualitative manner. We have invariably used a susceptibility ζ=1 for exotic cattle in this work. Therefore, a homogeneous network can be considered as a network of only exotic cattle (susceptibility ζ=1).

Simulations are performed for a variety of initial outbreak conditions, such as a single location versus multiple location outbreaks with varying cattle populations, infection rates, and cattle movement probabilities. We configure the model to investigate and quantitatively evaluate the relative impacts of mosquito control, livestock movement regulations, and diversity in cattle populations. We explore different simulation sets, each consisting of a number

of simulation scenarios. For each scenario, we performed 100 simulations.

We present simulation results for different values of k as well as two ranges of infection rate β. We investigate the number of cows in different compartments in the *SEIR* model by choosing a set of values of β (0.001, 0.005, 0.01, and 0.03), and starting with an infected cow in the Kabale municipality for each simulation. We choose a medium cattle movement probability constant β=0.01 to reduce the number of simulations. We also conduct simulations with different locations for the initially infected cattle, as well as single-location and simultaneous multiple-location RVFV epizootic outbreaks. We configure the network with values of k=0.01 and performed simulations for β=0.001, 0.005, 0.01, and 0.03 to reduce the number of simulation scenarios.

2.3 Numerical simulations results and discussion

2.3.1 Simulation set I

In this set, simulations are initiated with a single infected cow in the Kabale municipality and three values of k (0.001, 0.01, and 0.1), two ranges of β (0.0001-0.005 or 0.001-0.048), and two network topologies (homogeneous and heterogeneous), producing four scenarios:

- **Scenario 1**: Homogeneous network and β range 0.0001-0.005
- **Scenario 2**: Homogeneous network and β range 0.001-0.048
- **Scenario 3**: Heterogeneous network and β range 0.0001-0.005
- **Scenario 4**: Heterogeneous network and β range 0.001-0.048

Scenario 1

The simulation for β ranging between 0.0001-0.005 (the lower range) is presented in Figure 2.3 for three different values of the exponential constant k and a homogeneous network. We perform simulations for 100 days and recorded the fraction of infected cattle for each value of β.

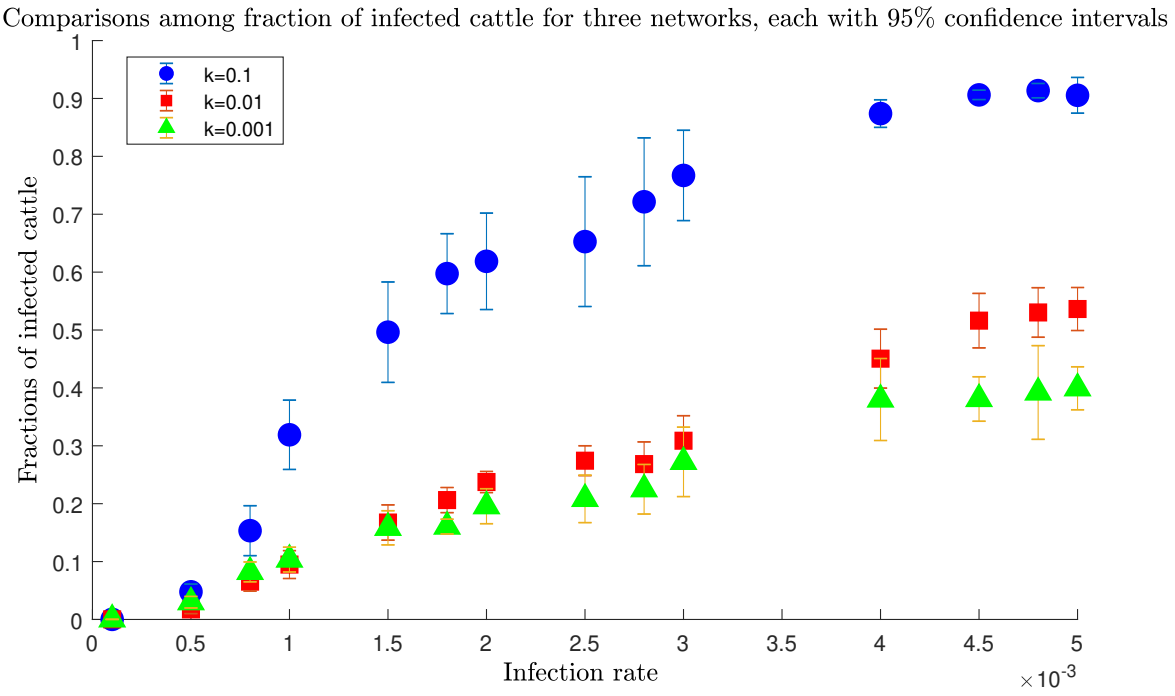


Figure 2.3: *Comparisons among fractions of infected cattle for a homogeneous network, three different values of k, and lower range of β; blue dots shows the fraction of infected for k=0.1, while red rectangles and green triangles show the fraction of infected for k=0.01 and 0.001 respectively. For the same value of infection rate, we always have more infected cattle for greater values of k (0.1) than the smaller ones (0.01 and 0.001). Therefore, increasing movement probability means more widespread epizootic. For example, the fraction of infected cattle at β=0,005 was 0.399, 0.537, and 1 for k=0.001, 0.01, and 0.1, respectively.*

From Figure 2.3, for k=0.01 and 0.001 it is evident that for β=0.005, the infection reaches half the population after 100 days. However, almost all of the cows are infected in the network after 100 days for k=0.1. A value of k=0.1 means extensive cattle movement between locations, which makes the whole network infected. Therefore, network structure plays a prominent role in RVFV spreading when the value of β is small.

Scenario 2

In the second set of simulations, we use a β ranging from 0.001 to 0.048 for the homogeneous network. Simulation results using these values are presented in Figure 2.4 for all three parameter values of k.

From Figure 2.4, it is indicative that the full network becomes infected very quickly for

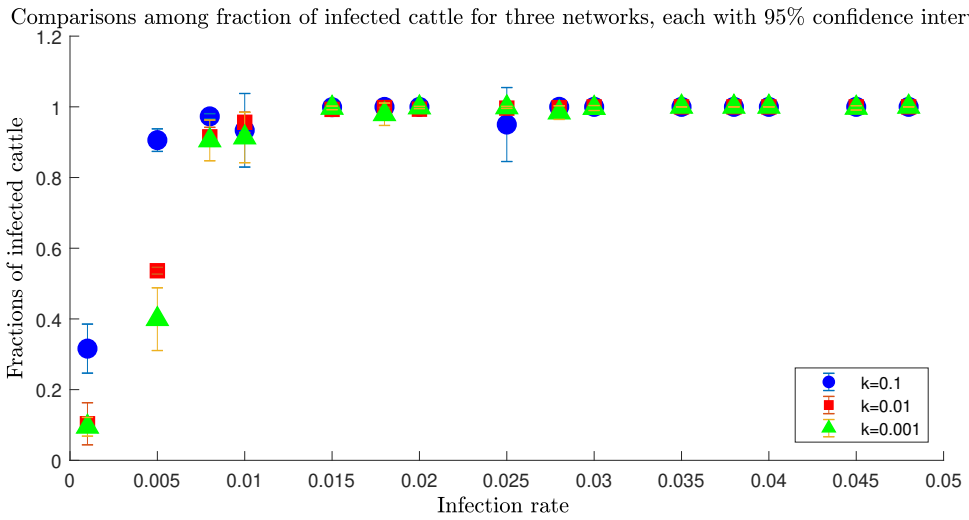


Figure 2.4: *Comparisons among fractions of infected cows for the homogeneous network, three different values of k, and upper range of β; fractions of infected for all three values of k are almost overlapping, therefore, not sensitive to the movement probability. They reach a value very close to one, i.e., the whole network becomes infected when infection rate β reached 0.01 for the three networks. Therefore, fractions of infected cows were also independent of the infection rate.*

this particular range of β irrespective of movement probabilities (k). Therefore, for a higher abundance of mosquitoes and favorable weather conditions, the spread of infection does not depend on the network structure and spread throughout the whole network very quickly.

Scenario 3

In this scenario, we repeat simulations for the heterogeneous network and lower β range and presented simulation results in Figure 2.5. An increasing trend is observed in the fractions of infected cows with an increase of β and k. For $k = 0.001$ and 0.01, there is little difference; however, for k=0.1 the increase of the infected fraction was faster with increasing β. Therefore, cattle movement needs to be reduced during an epidemic outbreak.

Scenario 4

Simulation results for the heterogeneous network, and the upper range of β are shown in Figure 2.6. For all three values of k, the fraction of infected cows reached one very quickly,

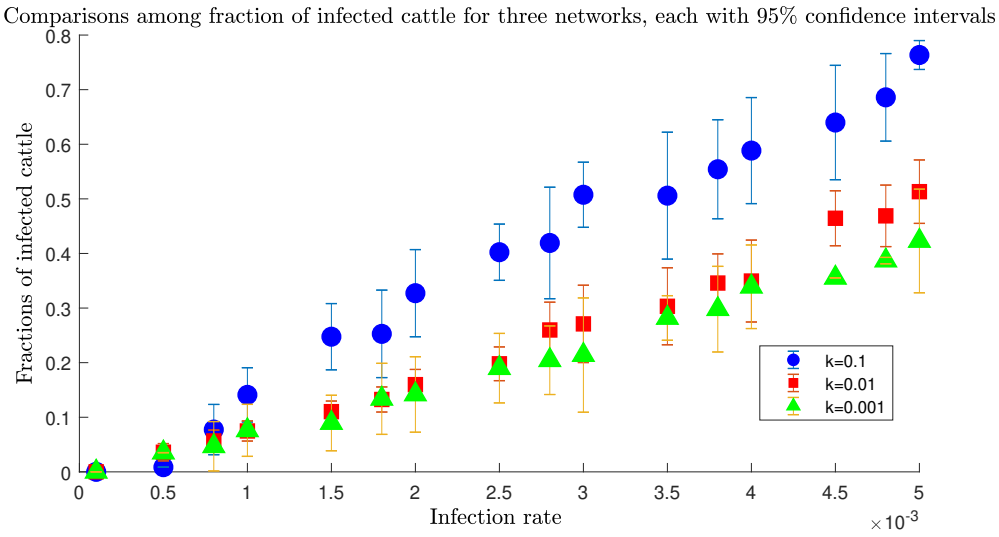


Figure 2.5: *Comparisons among fractions of infected cows for the heterogeneous network for three different values of k and lower range of β; for k=0.001 and 0.01, the maximum fraction of infected cows is less than 0.5 for the highest value of infection rate in the lower range. This means that after the simulation period, half of the cows become infected. However, for k=0.1, the fractions of the infected cow reach up to 0.8. Therefore, we needed to reduce the value of k, i.e., cattle movement, to reduce the fraction of infected cows.*

near a β value of 0.03. After that, all cows became infected regardless of the values of β and k.

Trends of the fraction of infected cows for both homogeneous and heterogeneous networks are similar in both lower and upper ranges of β. However, differences exist between fractions of infected cattle from homogeneous compared to heterogeneous networks for the same value of k and the same range of infection rate values. Comparisons between fractions of infected cows for homogeneous and heterogeneous networks are shown in Figure 2.7 and Figure 2.8. Figure 2.7 shows comparisons between fractions of infected cows for homogeneous and heterogeneous networks for the lower range of β values. It shows that the homogeneous network has more infected cows for the same values of β compared to the heterogeneous network. Lesser susceptibility of indigenous cows results in fewer infections among them. Since we specify that indigenous cattle are less susceptible to infection, the heterogeneous network results in fewer infected cattle than the homogeneous network where all cattle are exotic and highly susceptible.

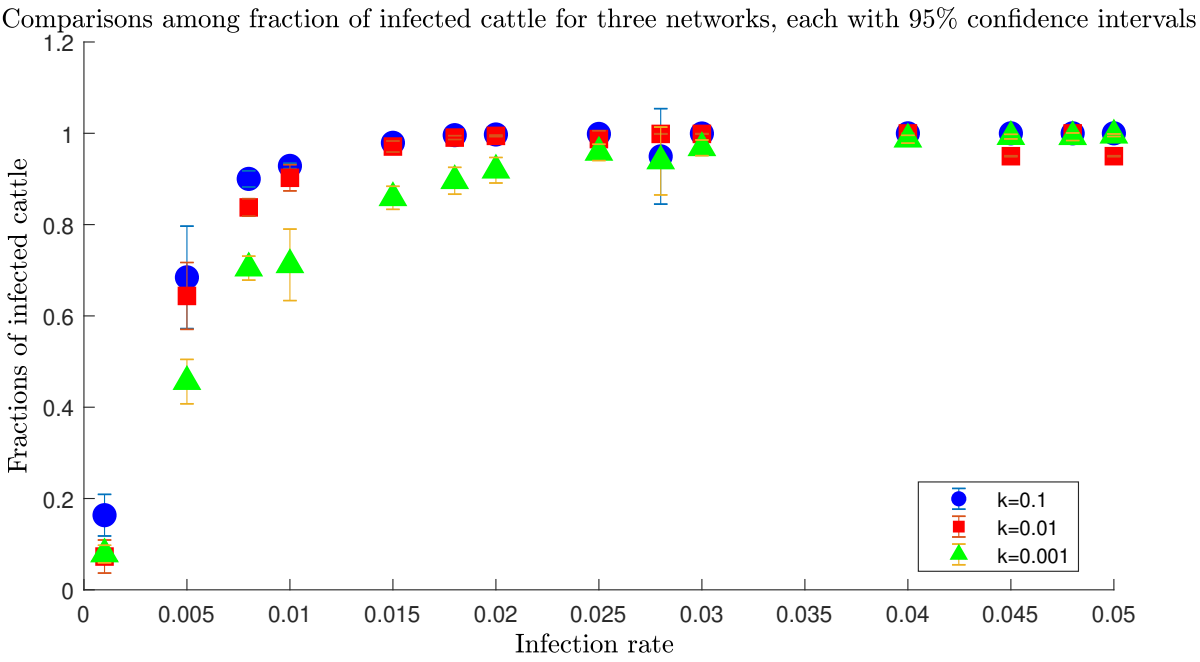


Figure 2.6: *Comparisons among fractions of infected cows for the heterogeneous network for three different values of k and upper range of β; the fractions of infected reaches towards one rapidly and when the value of infection rate is 0.03, the fraction of infected become one for all three networks.*

In Figure 2.8, the difference between the fractions of infected cows for two networks is negligible when β >0.005. Therefore, lesser susceptibility of indigenous cows cannot compensate for the higher mosquito abundance and results in a similar infection spreading in both homogeneous and heterogeneous networks.

Comparisons between homogeneous and heterogeneous networks show reduced susceptibility of indigenous cattle means fewer infected cows for lower mosquito abundance during an RVFV epizootic. Therefore, higher proportions of indigenous cows across locations would have the potential to reduce the numbers of infected cows and thus produce a more contained epizootic. In summary, simulations with lower infection rates result in increased fractions of infected cows with increasing movement probability. However, for high infection rates, the fraction reached one, and there is little difference in infected cattle fractions while increasing movement probabilities. From these observations, we conclude that, for low infection rates (low mosquito abundance), restricted cattle movement will reduce the

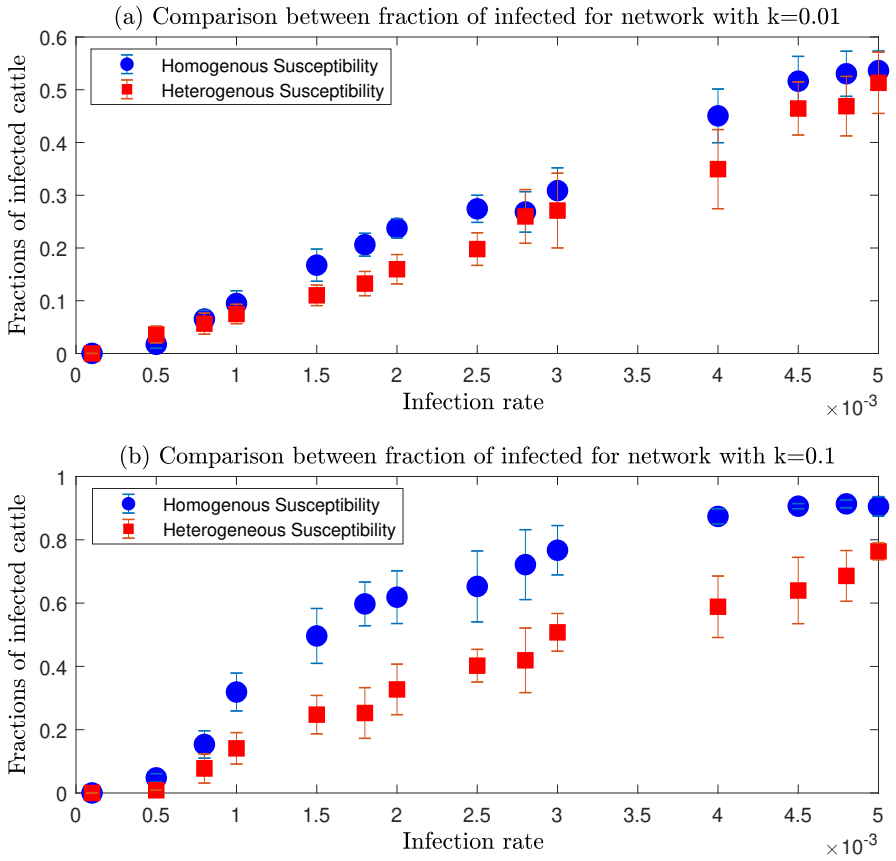


Figure 2.7: *Comparisons among fractions of infected cows for heterogeneous and homogeneous networks for lower range of β and a) k=0.01 and b) k=0.1.*

number of infected cows. Higher infection rates infect the whole network, regardless of cattle movement probability or mosquito abundance /infection rate. Therefore, for a period of low mosquito abundance, cattle movement should be restricted to contain the epizootic to a minimum level; whereas, periods of high mosquito abundance (high infection rates) would require both mosquito control and cattle movement restriction. Comparisons between fractions of infected for homogeneous versus heterogeneous networks suggest that diversity in the network resulted in fewer infected cows for similar values of infection rates and cattle movements.

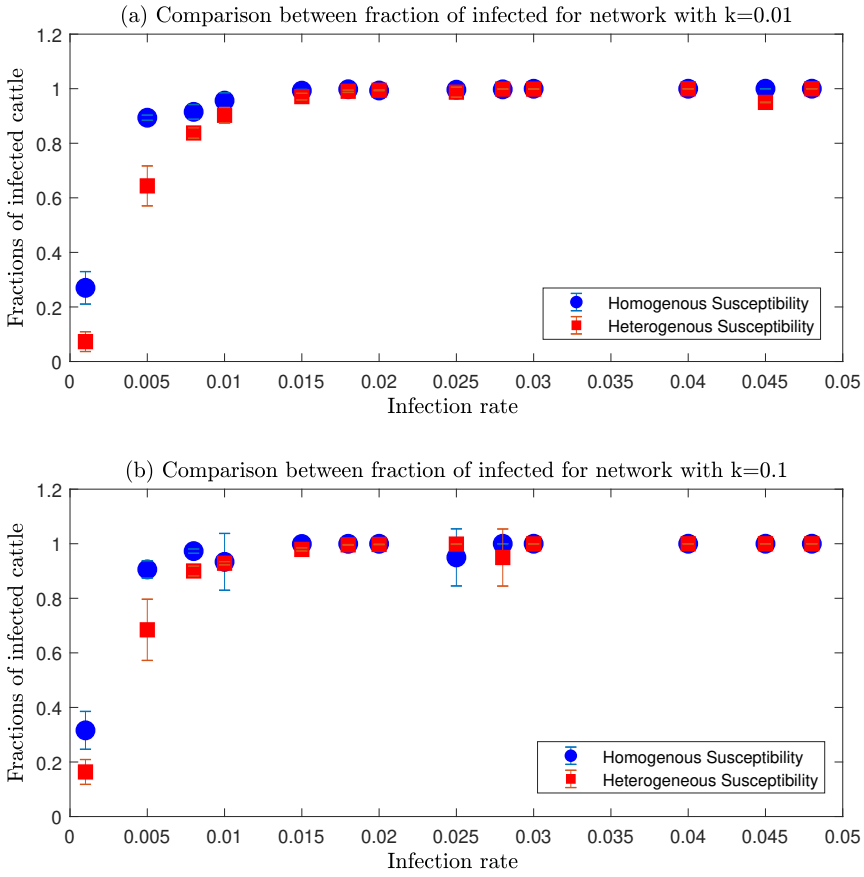


Figure 2.8: *Comparisons among fractions of infected cows for heterogeneous and homogeneous networks for the upper range of β and a) k=0.01 and b) k=0.1.*

2.3.2 Simulation set II

Simulations are conducted starting with a single infected cow in the Kabale Municipality, using both homogeneous and heterogeneous networks with k=0.01 and for β= 0.001, 0.005, 0.01, and 0.03 for each network, and produced two scenarios:

- Scenario 1: Homogeneous network
- Scenario 2: Heterogeneous network

For each scenario, we assume four different β to represent the entire range of infection rates used in the simulation set I. Instead of using different movement probability constants

(k=0.001, 0.01, and 0.1) we choose k=0.01 for both homogeneous and heterogeneous networks.

Scenario 1

Simulation results for homogeneous network and single infected cow in the Kabale municipality are presented in Figure 2.9. As β increase from 0.001 to 0.03, fractions of recovered reached one very quickly. It is worth noting that "fractions of recovered" means these were the cows that had been infected in the first place. All infected cows moved to the recovered compartment as we had not considered any disease-induced mortality in the model. Therefore, the fraction of recovered cows was considered the cumulative fraction of infected cows for our specific model.

Table 2.2: *Table shows maximum infected fractions of cows, peak infection time, and rate at which that maximum is attained for a homogeneous network.*

Infection rate β	Maximum Infected Fraction	Peak Infection Time	Rate
0.001	0.0095	45	2.1268e-04
0.005	0.065	87	6.919e-04
0.01	0.0806	64	0.0013
0.03	0.1345	31	0.0043

Scenario 2

Simulation results for a heterogeneous network with the initial condition of a single infected cow in the Kabale municipality are presented in Figure 2.10.

Table 2.3: *Table shows maximum infected fractions of cows, peak infection time, and rate at which that maximum is attained for a heterogeneous network and a single infected cow in the Kabale municipality.*

Infection rate β	Maximum Infected Fraction	Peak Infection Time	Rate
0.001	0.0056	60	9.333e-05
0.005	0.0365	100	3.6479e-04
0.01	0.0739	76	9.7690e-04
0.03	0.1181	43	0.0027

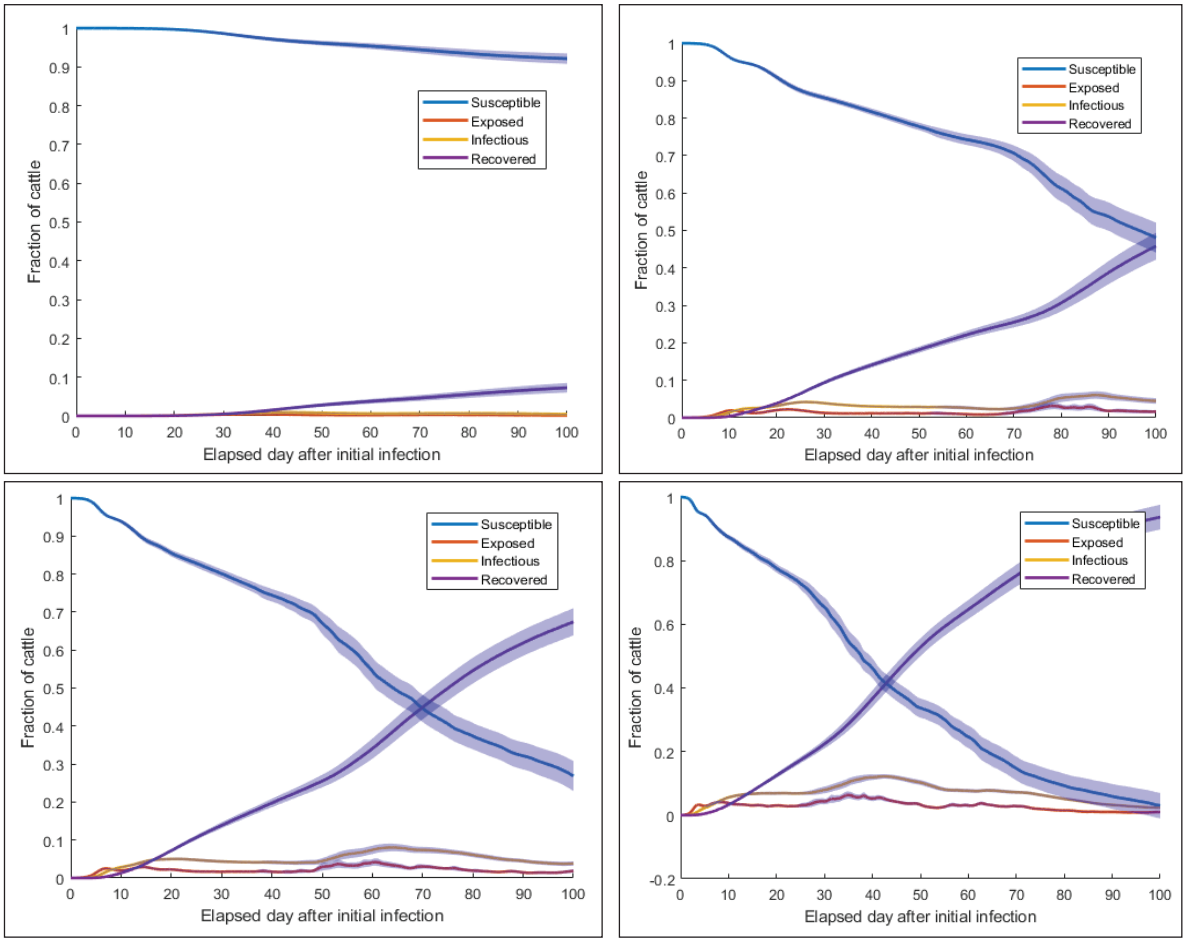


Figure 2.9: *Fraction of cows in each compartment with a 95 percent confidence interval for β=0.001 (top left), 0.005 (top right), 0.01 (bottom left), and 0.03 (bottom right) and for a homogeneous network; increasing β shows an increasing trend in the overall fractions of infected (cumulative fractions of recovered).*

Tables 2.2 and 2.3 shows that the rate at which the fraction of infected reached the maximum is increased with increasing β. However, a trend appears that when the value of β is minimal, i.e., β <0.005, the fraction of infected reached the maximum faster for both homogeneous and heterogeneous networks than higher values of β (β >0.005). This faster increase can be attributed to the fact that, when the value of β is minimal, the infection takes a long time to reach distant locations. Therefore, cows in the Kabale Municipality becomes infected within our simulation period of 100 days. The infection does not reach distant locations, reinforcing the impact of reduced vectorial capacity in containing the outbreak. The infection reaches distant locations at a slower rate than the rate of infecting local animals

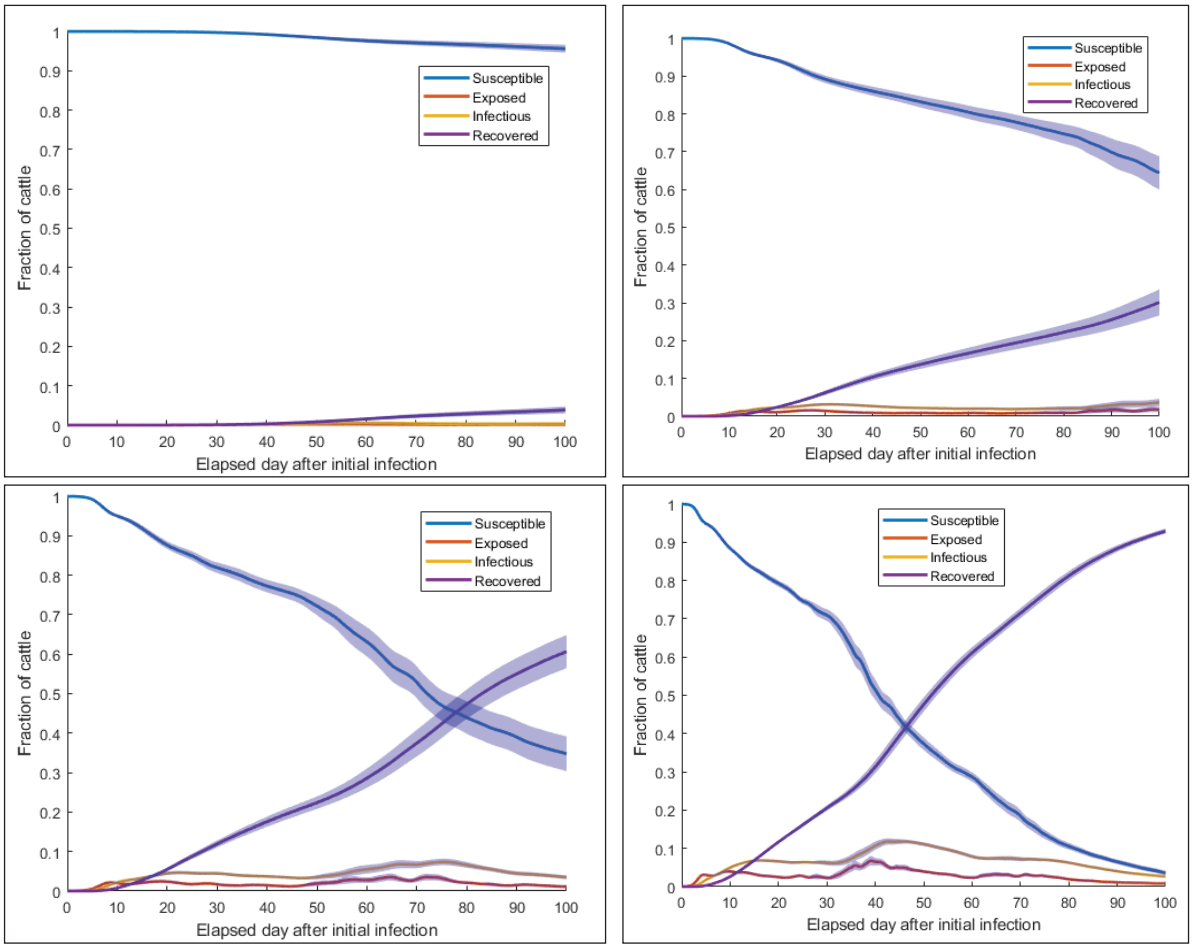


Figure 2.10: *Fraction of cows in each compartment with a 95 percent confidence interval for β=0.001 (top left), 0.005 (top right), 0.01 (bottom left), and 0.03 (bottom right) and for heterogeneous network; increasing β shows an increasing trend in the overall fractions of recovered (cumulative fractions of infected) which reaches to almost one for β=0.03.*

with increasing β. However, when the infection reaches distant locations, higher numbers of infected cows appear in the network. This is evident from the maximum fraction of cows, which is higher than the maximum fraction of infected cows for β=0.001. However, when β is significant (0.03), the time to reach the maximum is less than the time taken for β=0.001. Simulation results indicate that an increase in the infection rate expedited the spread of the epizootic in distant locations as well as the number of infected cows. Therefore, mosquito control was crucial to contain the epizootic in the initial outbreak location while taking proper measures to care for infected cows.

2.3.3 Simulation set III

This simulation set consisted of the following four scenarios:

- Scenario 1: Infection starting at a single location (Bubale) with the maximum number of cows
- Scenario 2: Infection starting simultaneously at three locations (Bauble, Rubaya, and Hamurwa) with the maximum number of cows
- Scenario 3: Infection starting at a single location (Muhanga T/C) with a minimum number of cows
- Scenario 4: Infection starting simultaneously at three locations (Bukinda, Muhanga, and Ruhija) with the minimum number of cows

The time to reach maximum infection for each scenario with infection rate β is shown in Figure 2.11, which shows a summary of Scenario 1 and 2 simulations when the initial RVF outbreak occurred in a single location or simultaneously at multiple locations, respectively.

The time required to reach maximum infection is shorter for simultaneous outbreaks, regardless of the network structure than for single-location outbreaks for similar values of the infection rate β(Figure 2.11(a)). The spreading of infection through the network is slower in the heterogeneous network for both single and simultaneous outbreaks. Infections spread slowly for the single-location outbreak in the network compared to the rate of spread in simultaneous outbreaks, which is reflected by the higher peak incidence time. For β=0.001, peak infection time was close to 100 days for all simulations except simultaneous outbreaks in homogeneous networks (Figure 2.11(a)). This means the peak is not attained, and the number of infected cattle was exponentially increasing. When β is increased to 0.01 (high mosquito abundance), the time to reach the peak is reduced drastically for all single and simultaneous outbreaks.

When the value of the infection rate increased to 0.005-0.03, there is a correlated decrease in peak incidence time as the whole cattle network becomes infected very quickly (well before

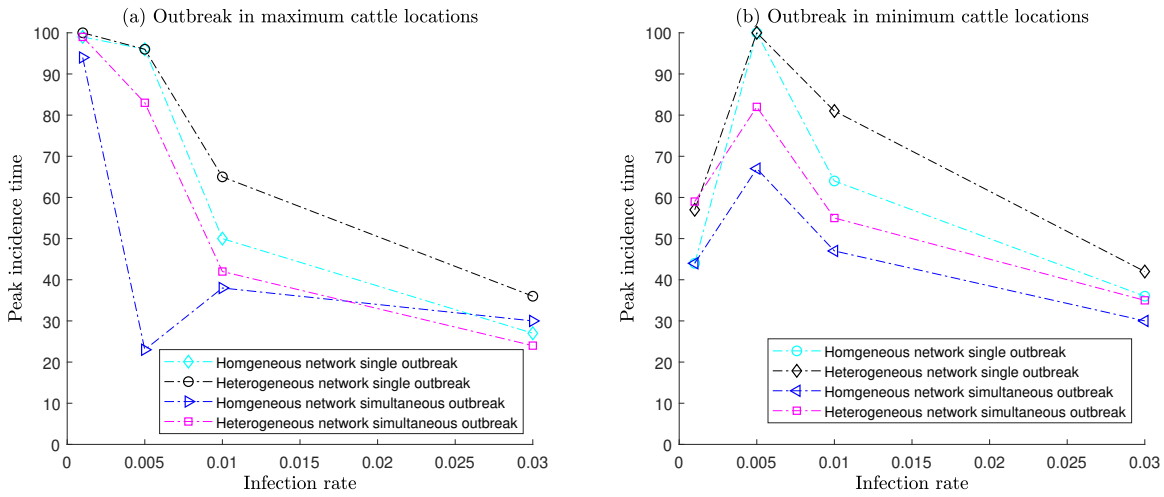


Figure 2.11: *Peak infection time with infection rate and for outbreaks starting in location/locations with (a) higher number of cows and (b) fewer number of cows.*

a 100-day simulation period) irrespective of outbreak location(s). Outbreaks in locations with a higher number of cows result in simultaneous virus introduction to distant locations having numerous connections. Therefore, with the increase in mosquito abundance, peak infection time decrease accordingly.

Figure 2.11(b) represents peak incidence time when the RVF outbreak occurred in location(s) with fewer cows than other locations. For lower mosquito abundance (β=0.001), the infection does not reach distant locations. Instead, it was quickly confined to the initial location(s), as evident from smaller values of the peak infection time. However, with increasing β, the peak time returns to its regular pattern, shown in Figure 2.11(a).

2.4 Conclusions

Recent RVFV activity in Uganda demonstrated the capability of the virus to spread into new regions through livestock movements and underscored the need to develop effective mitigation strategies to reduce transmission and prevent spread among cattle populations. We simulated RVFV transmission among cows in 22 different locations of the Kabale District in Uganda using real-world livestock data in a network-based model. This model considered livestock as a spatially explicit factor in different locations subjected to specific vector and environmental

factors. It was configured to investigate and quantitatively evaluate the relative impacts of mosquito control, livestock movement, and diversity in cattle populations on the spread of the RVF epizootic. We concluded that cattle movement should be restricted for periods of high mosquito abundance to control epizootic spreading among locations during an RVF outbreak. Importantly, cattle populations with heterogeneous genetic diversity as crossbreeds were less susceptible to infection compared to homogeneous cattle populations. Given the same initial conditions, the heterogeneous cattle population is less susceptible to infection than the homogeneous population. Simultaneous outbreaks in different locations will result in more infected cows at a faster rate of spreading compared to a single-location initial outbreak.

Chapter 3

Estimation of parameters and basic reproductive ratio using Sequential Monte Carlo Filter[1]

3.1 Introduction

Japanese encephalitis (JE) is a vector-borne viral disease endemic to the large part of Asia and the Pacific. It causes significant mortality among the human population every year. Therefore, it became an important public health concern[65]. Infectious disease mathematical models are developed to simulate outbreaks and empirically test possible mitigation strategies that reduce transmission and prevent spreading[66]. The complex transmission process of Japanese encephalitis can be expressed as a mathematical model– quantifying the interactions between the host/vector species in different infectious states. Mathematical JE model plays a very important role in a deeper understanding of the dynamics of the pathogen spread and hence can contribute to the reduction or even completely stop the progression of the spreading.

Mathematical disease models can be very simple compartmental models or more com-

[1] This chapter is a reformatted and slightly modified version of our published article[64], Copyright © 2017, IEEE

plex ones when required according to the specific disease[67]. Once an appropriate model is formulated, that can be combined with different statistical or control systems framework to acquire information about the disease spread and forecasting. A mathematical model of JE should include environmental factors to accurately represent the biological transmission cycle.

Mukhopaddhay et al. (1994)[68], Ghosh et al. (1999)[4] and Naresh et al. (2009)[69] proposed some important models. The authors divided concerned populations (vector, host, or reservoir) in different compartments and expressed the epidemic dynamic with some differential equations. The transition from one compartment to another depends upon some transfer rates, which are expressed as parameters, dependent mainly on the biology of the individuals and some other concerned species (mosquito, and birds) and their density and contact pattern. The success of these models depends crucially on the proper estimation of parameters[65]. In all these models, the authors find an expression for calculating the basic reproductive ratio. The basic reproductive ratio indicates the number of secondary cases evolved from one single infectious case[70]. A very simple mitigation strategy for any epidemic is concerned with this ratio. If we can reduce the basic reproductive ratio below one, we can successfully prevent a disease from spreading. Therefore, a proper estimation of the basic reproductive ratio is mandatory for the successful mitigation of an epidemic. The reproductive ratio is a quantitative measure of the severity of the disease to the public health authorities[71].

All models mentioned above are deterministic, formulated based on simplified assumptions about the actual stochastic spreading process. Riad et al.[65] proposed an individual level stochastic model. Authors develop a model for a scenario of JE epidemiology in the United States. Only one population—feral pigs in three spatial locations—is represented at the individual animal level via a connected network and suggested mitigation strategy from their simulations. However, all these models, whether deterministic or stochastic, are crucially dependent on the estimation of parameters. The basic reproductive ratio is also dependent on the model parameters. Therefore, we focus on estimating the parameters for spreading JE from real-life incidence data.

In this chapter, we formulate a *susceptible-exposed-infected-recovered (SEIR)* model for the JE progression and use Sequential Monte Carlo (SMC) filter, an online inference method to simultaneously estimate the disease state and parameters as new data about the disease are available. In SMC, we iteratively sample from the posterior distribution of parameters until the parameters converge to stationary values. As an example, we use monthly incidence data of JE in the Philippines to estimate associated parameters with time. Sequential Monte Carlo filter (particle filter) is particularly important for its inference to the nonlinear and uncertain epidemic models. This method will allow us to estimate the parameters associated with the *SEIR* model, which will eventually allow us to estimate the basic reproductive ratio of the JE epidemic.

3.2 Background

3.2.1 Japanese encephalitis compartmental Model

In this chapter, we assumed a *Susceptible-Exposed-Infected-Recovered (SEIR)* model for JE transmission in the Philippines. The model can be expresses as the following difference equation-

$$S_{t+1} = S_t - \beta S_t I_t \tag{3.1}$$

$$E_{t+1} = E_t + \beta S_t I_t - \delta E_t \tag{3.2}$$

$$I_{t+1} = I_t + \delta E_t - \gamma I_t \tag{3.3}$$

$$R_{t+1} = R_t + \gamma I_t \tag{3.4}$$

In these equations, *S, E, I*, and R are called the states. Each individual human can be in one of these four states. We have three parameters in this *SEIR* model- infection rate β, incubation rate δ and the recovery rate γ.

In this compartmental *SEIR* model, the size of host population is assumed to remain constant throughout the evolution time, i.e., $P = S+E+I+R$, and demographic effects are

ignored. For our particular *SEIR* model, the basic reproductive ratio is- $R_0 = \beta/\gamma$.

Many problems in science require estimation of the state of a system that changes over time (time-series) using a sequence of noisy measurements made on the system[34]. The state-space approach to time-series modeling focuses attention on the state vector of a system. The state vector contains all relevant information required to describe the system under investigation[34]. The state-space approach is convenient for handling multivariate data and nonlinear/non-Gaussian processes, and it provides a significant advantage over traditional time-series techniques for these problems[34].

In order to analyze and make inference about a dynamic system, at least two models are required: first, a model describing the evolution of the state with time (the system model) and, a second model relating the noisy measurements to the state (the measurement model)[34].

3.2.2 State-space model

State-space models are usually used for analysis of dynamic data and general statistical model. They are being increasingly used in epidemiology. The state-space model has two parts— observation equation and state evolution equation shown in equations (3.5) and (3.6), respectively.

$$y_k \sim p_{y,k}(y_k|x_k, \theta) \tag{3.5}$$

$$x_k \sim p_{x,k}(x_k|x_{k-1}, \theta) \tag{3.6}$$

In these equations, x_k is the system state at time k, and y_k is observational data, and the θ denote the state-space model parameters. When data are collected sequentially, current state and parameters distribution is then conditional on the data observed up to that time. This distribution incorporates all state and parameter information up to time step k . It can be updated recursively using Bayes' rule:

$$p(x_k, \theta|y_{1:k}) \propto p(y_k|x_k, \theta)p(x_k, \theta|y_{1:k-1}) \tag{3.7}$$

Most of the time, the value of $p(x_k, \theta|y_{1:k})$ cannot be calculated analytically, and we need to use numerical methods such as extended Kalman filter, Gaussian Sum filter or Monte Carlo versions (particle filter).

3.2.3 Particle filter

Particle filtering is an SMC inferential technique based on repeated use of importance sampling. It aims to approximate the filtered distribution at time k through a weighted Monte Carlo realization from this distribution in terms of I particles, i.e.

$$p(x_k, \theta|y_{1:k}) \approx \sum_{i=1}^{I} w_i \delta(x_k^{(i)}, \theta_k^{(i)}) \tag{3.8}$$

In equation (3.8), I is the number of particle, w_i is the weight of that i^{th} particle and $\sum_{i=1}^{I} w_i = 1$. $(x_k^{(i)}, \theta_k^{(i)})$ is the i^{th} particle location and δ is the delta function. Among, all particle filters, the two basic ones are Auxiliary and Bootstrap particle filter. In these two particle filters, we assumed that the distribution of parameters are known. Therefore, in the description of these two filters, the parameters are not included.

3.2.4 Bootstrap particle filter

This is first successful particle filter. If we know the filtered distribution at time step k for the equation (3.9), then we can find the distribution in time k+1 using the following step-

$$p(x_k, \theta|y_{1:k}) \approx \sum_{i=1}^{I} w_i \delta(x_k^{(i)}, \theta_k^{(i)}) \tag{3.9}$$

1) Sample an index $j \in 1....i....I$ with associated probabilities $w_k^1, ...w_k^i, ...w_k^I$

2) Sample $x_{k+1}^{(j)} \sim p(x_{k+1}|x_k^{(j)})$, and

3) Calculation of weights and renormalization:
$\bar{w}_{k+1}^{(i)} = p(y_{k+1}|x_{k+1}^{(i)})$ and $w_{k+1}^{(i)} = \bar{w}_{k+1}^{(i)} / \sum_{i=1}^{I} \bar{w}_{t+1}^{(i)}$

3.2.5 Auxiliary particle filter

The main problem with the bootstrap particle filter is the small value of the weight when $p(y_k|x_k^{(i)})$, which will reduce the effect of these particles in the approximation of $p(x_k|y_k)$. In the auxiliary particle filter, this problem is rectified using auxiliary weights. The following steps are followed for auxiliary particle filter-

Step 1: Point estimate calculation: $\mu_{k+1}^{(i)} = E(x_{k+1}|x_k^i)$

Step 2: Auxiliary weights calculation:

$\bar{g}_{k+1}^i = w_k^{(i)} p(y_{k+1}|\mu_{k+1}^{(i)})$ and $g_{k+1}^{(i)} = g_{k+1}^{(i)} / \sum_{i=1}^{I} g_{k+1}^{(i)}$

Step 3: a) Sample an index $j \in 1....i....I$ with associated probabilities $g_{k+1}^1, ...g_{k+1}^i, ...g_{k+1}^I$

b) Sampling $x_{k+1}^{(i)} \sim p(x_{k+1}|x_k^{(j)})$

c) Calculation of weights and re-normalization:

$w_{k+1}^{(i)} = \frac{p(y_{k+1}|x_{k+1}^{(i)})}{p(y_{k+1}|\mu_{k+1}^{(j)})}$ and $w_{k+1}^{(i)} = w_{k+1}^{(i)} / \sum_{i=1}^{I} w_{k+1}^{(i)}$

The point estimates above can be the mean or any other point estimate. In order to simultaneously estimate the time-evolving states and fixed parameters using either the BF or APF, it is necessary to incorporate the fixed parameters into the state with degenerate evolutions[72;73].

However, these aforementioned particle filters are prone to degeneracy due to the use of fixed parameter values while using only the state variables as the primary concern. Liu and West[74] builds a particle filter based on the auxiliary particle filter and provides a general way of fighting degeneracy in fixed parameters. This particle filter is called the kernel density particle filter.

3.2.6 Kernel density particle filter

Kernel Density particle filter has advantages over other particle filters with respect to its robustness against degeneracy as no fixed parameter set are used here. The Kernel density particle filter is applied to update and estimate $p(x_{k+1}, \theta|y_{1:k+1})$. At the initial time step $k = 1$, weights for all particles are equal to I^{-1}, and initial state and parameters are generated by random sampling from a prior probability density functions $p(\theta_0)$ and $p(x_0)$[71;72].

We present the steps followed for this filtering method to estimate the posterior distribution in the following section.

The algorithm is following-

Step 1: $m_{k+1}^{(i)} = \alpha\theta_k^{(i)} + (1-\alpha)\bar{\theta}_k$

$\bar{\theta}$ is a weighted sample mean and α is a parameter that controls the smoothness of the estimation. The value of α has a complex structure $\alpha = 1 - h^2$ and where h is expressed as $1 - (\frac{3\xi-1}{2\xi})^2$ where $\xi \in (0,1)$[74].

Step 2: Mean Calculation: $x_{k+1}^{(i)}$: $\mu_{k+1}^{(i)} = E(x_{k+1}|x_k^{(i)}, \theta_k^{(i)})$

Step 3: Calculation of auxiliary weights and re-normalization:

$$g_{k+1}^{i} = w_k^{(i)} p(y_{k+1}|\mu_{k+1}^{(i)});$$

$$g_{k+1}^{(i)} = \bar{g}_{k+1}^{(i)} / \sum_{i=1}^{I} \bar{g}_{k+1}^{(i)}$$

Step 4: Repetition for all particles $i = 1, 2..i..I$

(a) sample an index $j \in 1....i....I$ with associated probabilities $g_{k+1}^1, ...g_{k+1}^i, ...g_{k+1}^I$

(b) Parameter generation: $\theta_{k+1}^{(i)} \sim N(m_{k+1}^{(j)}, \sigma_{k+1}^p)$, σ_{k+1}^p is the weighted sample covariance for the parameters at each time step

(c) Sampling $x_{k+1}^{(i)} \sim p(x_{k+1}|x_k^{(j)}, \theta_{k+1}^{(i)})$

(d) Calculation of weights and re-normalization:

$$\bar{w}_{k+1}^{(i)} = \frac{p(y_{k+1}|x_{k+1}^{(i)}, \theta_{k+1}^{(i)})}{p(y_{k+1}|\mu_{k+1}^{(j)}, m_{k+1}^{(j)})}$$

$$w_{k+1}^{(i)} = \bar{w}_{k+1}^{(i)} / \sum_{i=1}^{I} \bar{w}_{k+1}^{(i)}$$

These particle filters are very efficient in parameter estimation and predicting future dynamics. Any epidemiological model[61] needs parameters to be estimated depending upon the region for which the model is proposed. The effectiveness of knowledge obtained from the epidemiological model is crucially dependent on the suitable parameter choice. Therefore, in this article, we will focus on parameter estimation using Kernel density particle filter from

the real incidence data. Although We propose this model for JE, however, our method has the ability to be adapted for any epidemic model.

3.3 Materials and method

3.3.1 Japanese encephalitis state-space model

We propose the following set of stochastic difference equations to capture the stochasticity in the spreading process–

$$E_{k+1} = E_k + \beta I_k - \delta E_k - \epsilon_\beta + \epsilon_\delta \tag{3.10}$$

$$I_{k+1} = I_k + \delta E_k - \gamma I_k - \epsilon_\delta + \epsilon_\gamma \tag{3.11}$$

$$R_{k+1} = R_k + \gamma I_k + \epsilon_\gamma \tag{3.12}$$

In the above equations, the ϵ_c where $c \in \beta, \delta, \gamma$, are random components which has a mean zero and variance c/P^2 where P is the total population in the Philippines. In these equations, we assumed the fraction of susceptible population remains at a constant value $S = 1$ as the total number of incidence is very small compared to the total population. We assumed that, our *SEIR* model follows a normal distribution. Therefore, the state-space model is–

$$(x_{k+1}|x_k, \theta) \sim N_\tau(\mu(x_k, \theta), \sigma(\theta)) \tag{3.13}$$

where N_τ represents truncated normal distribution of E_k, I_k and R_k.This distribution has the mean μ and the variance $\sigma(\theta)$. The mean and the variance are expressed as the following–

$$\mu(\mathbf{x_k}, \theta) = \begin{bmatrix} E_k + \beta I_k - \delta E_k \\ I_k + \delta E_k - \gamma I_k \\ R_k + \gamma I_k \end{bmatrix} \tag{3.14}$$

$$\sigma(\theta) = \frac{1}{P^2}\begin{bmatrix} \beta+\delta & -\delta & 0 \\ -\delta & \gamma+\delta & -\gamma \\ 0 & -\gamma & \gamma \end{bmatrix} \tag{3.15}$$

In the above matrices the, P=100000000 is the total population. We used the data set obtained from Lopez et al.[75], which consists of the monthly confirmed JE cases from different hospitals in the Philippines from March 2011-March, 2014. Although the total population of the Philippines is 100000000, the confirmed infected cases are proportionally small. For the observation equation of the state-space model of JE, we took the cumulative number of infected people and expressed it as a log-normal distribution with a mean $b_1(I_k + R_k)^{\zeta}$ and variance Σ. In particular, b_1 is a multiplicative constant, and ζ is a power-law exponent. We choose their value to be 1.05 and 0.95, respectively. The variance was selected as 0.00088. These values were selected according to Shahtori et al.[71] and Sheinsen et al.[72].

3.3.2 Application of kernel density particle filter

In this work, among all particle filters discussed in the background section, we used a kernel density particle filter as we are particularly interested in the parameter estimation. Moreover, it has advantages over other particle filters concerning its robustness against degeneracy[72].

We apply a kernel density particle filter, to update and estimate $p(x_{k+1}, \theta_{k+1}|y_{1:k+1})$. At the initial time step $k = 1$, weights for all particles are equal to I^{-1}, and θ_0 (initial parameters) and x_0 (initial states) are generated by random sampling from prior probability density functions $p(\theta_0)$ and $p(x_0)$. As we have discussed in the background section, $p(x_{k+1}, \theta|y_{1:k+1})$ when the $k + 1^{th}$ observation becomes available. The kernel density particle filter we have used in this article is an adaptation of that used in Sheinson at el.[72] and Shahtori et al.[71] for our particular case of JE in the Philippines.

The implementation steps for this particle filter is presented in the kernel density particle filter section. Particle filter setup is crucially dependent on the priors, which will remove the chance of overfitting the data[76]. Therefore caution and proper knowledge about the system

are required before defining the priors.

Priors for parameters $p(\theta_0)$

The transfer of JE pathogen depends on the mosquito's abundance, temperature, rainfall, humidity, and various factors related to host species[65]. However, the incubation period γ is 5-15 days for the human population, and the recovery period is γ has a range of values 3-7 days[77]. Therefore, we use suitable beta distribution for them as we know the variability of these parameters. However, to select the distribution of β and population in the exposed class, we use the mean squared error (MSE) technique with 110 different combinations of parameter and state priors to run the simulation. We use the combination that gives the minimum MSE. We use random distribution for the prior of infection rate β as $\mathcal{U}(0.25, 0.70)$.

Priors for states $p(x_0)$

We had a single infected individual at the beginning of the simulation in March 2011. Therefore, we used a fixed value for the prior of the I, and it consists of one individual in the infected state. However, the dataset does not contain any information about the recovered and exposed class. For the exposed population, we took eight people in the exposed class. This prior for an exposed class provides us the minimum squared error in combination with β specified earlier. We do not consider any recovered individual initially as they did not participate in the spreading process.

The data we use has the monthly new incidences of JE in the Philippines from Lopez *et al.*[75]; however, the data were collected from different hospitals all over the country. Therefore if there were some non-hospitalized cases, they are not reflected in the dataset[75]. We have used the cumulative number of infected. Therefore that reflects the total number of infected and recovered up to a specific time step k, beginning from March 2011.

3.4 Simulation results and discussion

The filter is formulated with the priors previously described and $I = 15000$ particles for the simulation of the posterior $p(x_k, \theta_k | y_{1:k})$. The filter is set up to estimate the total number of individuals in each month in each state, as well as the evolution of parameters with time. The case data set is 37 months of JE cases, and therefore 37 different values of each parameter and state are estimated using the filter. Our observation is the cumulative number of infected, which we plot over time to compare the filter results and the actual cases (Figure3.1). We have used a log-normal distribution for fitting the data using the filter; therefore, there is a bias that makes the estimation look exponential.

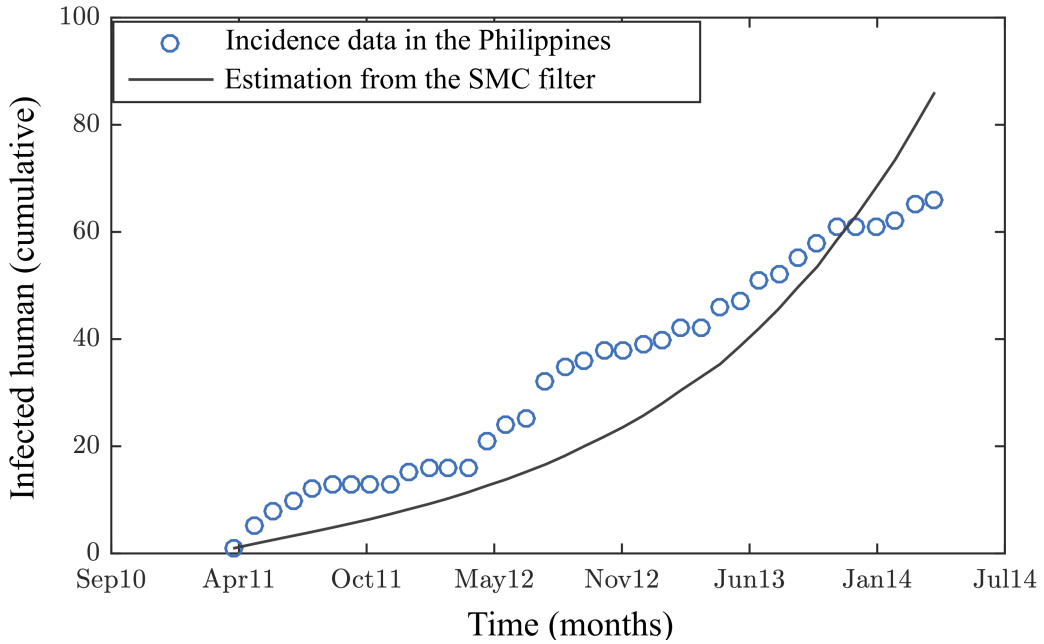


Figure 3.1: *Cumulative number of infected in the Philippines and their corresponding particle filter estimation.*

We have presented the variability of infection rate β with time in Figure 3.2 as this causes the number of infected to change seasonally and yearly. We can see several peaks and valleys in Figure 3.2, which takes into account the mosquito abundance in different seasons as well as in different years. For example, the peaks are always in the vicinity of the summer, while the valleys are in the winter. However, in our simulation results, there are

some deviations in the peak incidence period. We are using only hospitalized cases that have a high probability of being random. Therefore sometimes more cases may be reported to the hospital in winter than summer. This is responsible for deviant results from the seasonal pattern in JE incidences.

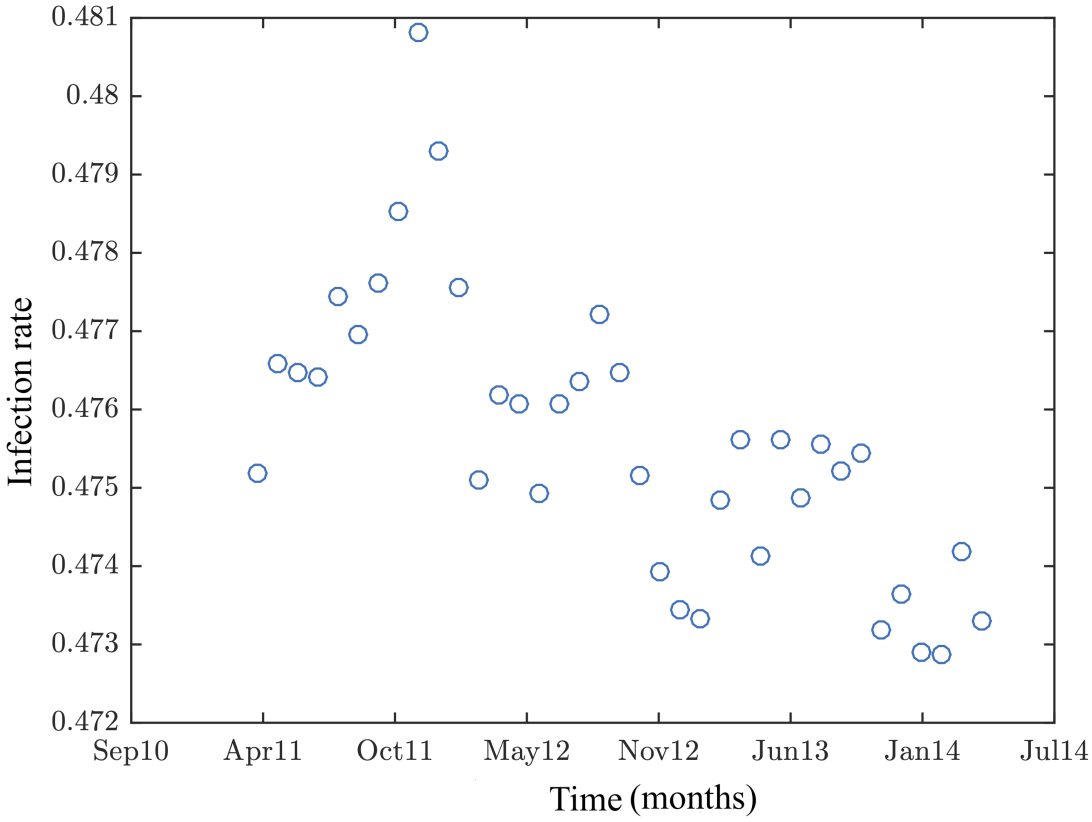


Figure 3.2: *Variability of the infection rate β with time.*

We have presented the basic reproductive ratio R_0 in Figure 3.3. Here, we can see that the value of R_0 is always higher than one. Therefore, the disease is present year-round, confirming the history of JE incidence in the Philippines and other tropical countries. Therefore, to stop the disease spread, control measures should be taken until the estimates of R_0 become less than one. Figure 3.2 and Figure 3.3 shows that the variability in the value of β and R_0 follow a similar trend. The reason behind this behavior is evident from the definition of R_0 as it is expressed as $\frac{\beta}{\gamma}$. The trend is almost similar because there is not much variability in the denominator value, recovery rate γ of the infected human.

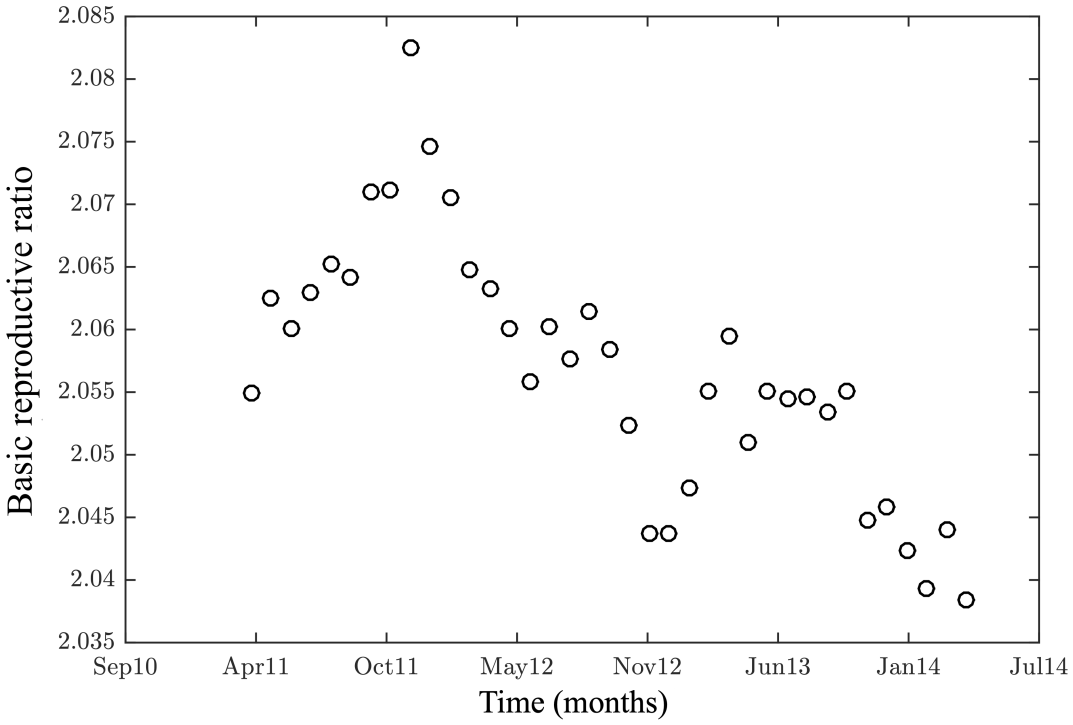


Figure 3.3: *Variability of the basic reproductive ratio R_0 with time.*

After estimating the parameters, we used the number of human cases concerning the first 32 months as the training set and performed predicted for the next five months. The prediction of the JE incidence for the estimated parameter set is shown in Figure 3.4 with the actual data.

The analysis of our results identifies that there is a seasonal variability of β as well as yearly variation (2011-2014). In Figure 3.2, we can see several peaks (April-August) and valleys, which indicates different years as well as different seasons. The peaks are always in the vicinity of summer (April-August) while the troughs are near winter (October-February). These peaks and valleys conform with the mosquito abundance during the corresponding periods. The temperature and humidity do not have significant fluctuations in the Philippines. This smaller fluctuation is reflected in β values, which varied between 0.471 and 0.482, a range of 0.011. The estimated basic reproductive ratio shows the value of 2.035-2.085. The value is always more than one, which explains that JE will remain endemic in the Philippines. Additionally, the calculated value of the R_0 is comparable with the R_0 value

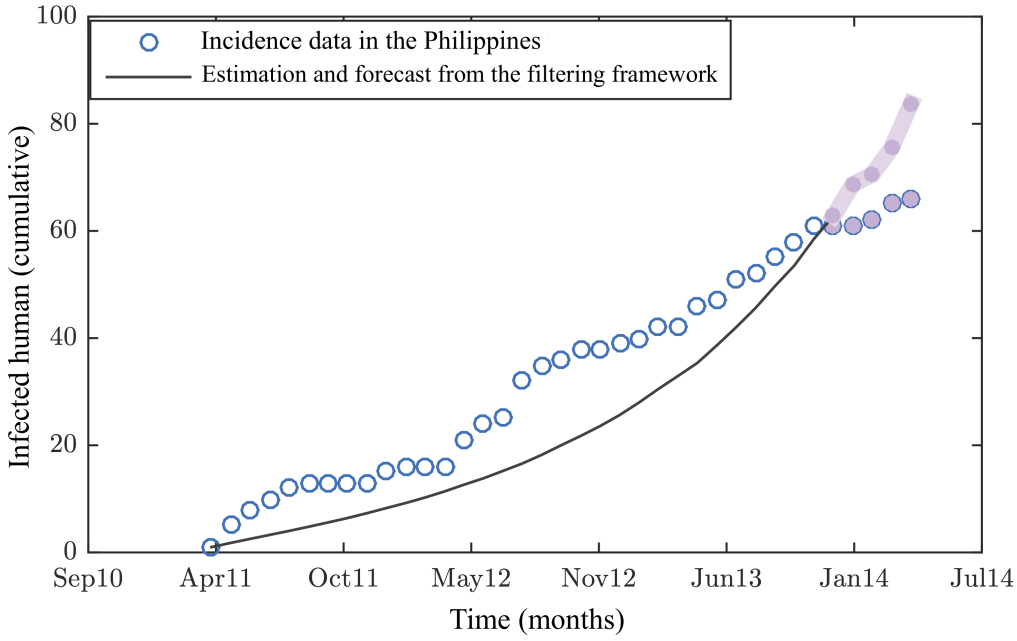


Figure 3.4: *Cumulative number of infected in the Philippines with the forecast for the last five data points. The data points reserved for the forecasting as well as the forecast is highlighted in pink.*

estimated for some diseases in the same *Flaviviridae* genre in countries with endemic cycles. For example, dengue in Pakistan has a value of 3.082[78], while the calculated value of R_0 for Zika has a value within the range 1.5-4.1 in the pacific islands (Yap, Micronesia, Tahiti and Moorea, French Polynesia and New Caledonia). All these countries have similar weather as the Philippines[79]. However, in temperate regions, the reproductive ratio is subject to much more variability as temperature and humidity extensively fluctuate seasonally. This explains the periodic outbreaks rather than year-long outbreaks. For example, in Italy, the estimated value of R_0 for West Nile Virus varied within 0.4-4.8 within a year, which is a closely related flavivirus and has a similar transmission cycle as JE[80]. However, the fluctuation in R_0 is not very pronounced for our simulations for the weather being always amiable to the spread of JE pathogen in the Philippines.

The most important control measure for any mosquito-borne disease is to reduce mosquito abundance by applying insecticides and making people aware of using personal measures to stay away from mosquito bites. Practical implementation and mathematical models have

shown that implementation of these control measures significantly reduces the infection rate β, which in turn reduces the basic reproductive ratio of R_0.

Khan et al.[78] calculated the basic reproductive ratio before and after using the control measure for dengue in Pakistan. The authors show that the value of R_0 changes from 3.0528 to 0.6293, becoming less than one. Therefore, a proper estimation of the reproductive ratio can help the public health administration take control measures against an epidemic. Our method of calculating the basic reproductive ratio at each time step once new incidence data are available can be one crucial predictor in rationing the limited resources in epidemic control. Once the application of control measures brings the R_0 below one, the use of resources can be reduced or withdrawn as the epidemic may die out soon.

3.5 Conclusions

We develop a sequential Monte Carlo filter to estimate the parameters in a stochastic model of infectious disease transmission. The network-based model's success crucially depends on the proper estimation of the parameters. This parameter estimation method is particularly important for its adaptability to the availability of new incidence data of any epidemic. As an application of the method, we apply this particular method for Japanese encephalitis transmission in the Philippines. Parameters estimated from the simulation of the particle filter shows seasonal as well as yearly variation. The evolution of the basic reproductive ratio is in compliance with the endemicity of JE in the Philippines. The state estimation from the estimated parameters shows a similar trend as the incidence data. Therefore, this framework is capable of estimating the realistic parameters for specific locations and infectious diseases, which can later be used for other purposes, such as investigating and quantitatively evaluate the relative impacts of different components in the network-based model.

Chapter 4

Short-term forecast and dual state-parameter estimation using ensemble Kalman filter[1]

4.1 Introduction

In most classical forecasting literature, constant parameters are used for the epidemic models. However, Shahtori et al. 2016 characterized and described temporal variation of the parameters for Ebola progression in Central Africa[71]. Although the parameters of an epidemic model can be estimated in a batch-processing scheme or from empirical data, it is not guaranteed that model behavior does not change with time[82]. Therefore, it is prudent to adjust the model with time, which can be accomplished by simultaneous estimation of parameters and states.

Several frameworks have been developed for the estimation of states and parameters for physical process models. Although the development of interactive state-parameter estimation using filtering methods is well established, their application to the field of epidemiology is relatively new, pertaining to a high level of non-linearity in the epidemic model. The

[1] This chapter is a reformatted and slightly modified version of our published article[81], Copyright © 2019, IEEE

original Kalman filter works for linear systems, while the extended Kalman filter uses the linear approximation of the non-linear systems. In 1994, Evenson proposed a Monte Carlo-based Kalman filter called ensemble Kalman filter (EnkF). The EnkF has initially been developed for dynamic state estimation and later adapted to be applied for parameter estimation[9;24;64;71–73].

This chapter focuses mainly on tailoring the dual state-parameter estimation method using EnKF to the vector-borne disease model with a specific data set of Japanese encephalitis (JE) from Taiwan. A *susceptible-exposed-infected-recovered (SEIR)* model of JE is used with the EnKF framework to simultaneous state-parameter estimation and for short- and mid-term forecasting. Once the forecasting is done, they are compared with the real incidence data to check the accuracy of the forecast. Forecasts for our specific data set are reasonable for the short distance in the future. However, long-term forecasts result in a loss of accuracy. We apply our filtering framework to forecast the epidemic to explore the efficacy of applying control measures during an epidemic. Simulation results show a reduction in mosquito abundance will significantly hinder the epidemic growth during an outbreak.

4.2 Materials and method

4.2.1 Sequential data assimilation and filtering

The primary purpose of the data assimilation is to characterize the state of a system at some future time from the information provided by the initial state. The states of an epidemiological system x_k, at time k is dependent on the observations up to time k. A variety of methods in estimation theory enable the recursive estimation of system state variables and the inference of model parameters[32]. Sequential data assimilation, also known as filtering methods, have been widely used in different engineering designs and forecasts. However, their use in epidemiological model development has only been started in the last two decades. The ability of an epidemiological model to make accurate predictions depends on the extent to which the model represents real-world transmission dynamics as well as the proper specification

of model parameters and initial conditions[33]. The initial condition is model state variable estimation at the start of a forecast. Filtering methods use the observations to recursively inform and train the model so that current conditions are better depicted and evolving outbreak characteristics (i.e., the trajectory of the epidemic curve) are better-matched[34]. The epidemic model can then be propagated into the future to make a more accurate and reliable forecast.

4.2.2 Kalman filter (KF)

Kalman filtering is an algorithm that uses a series of measurements observed over time, containing statistical noise and other inaccuracies. It produces estimates of unknown variables that tend to be more precise than those based on a single measurement alone, by using Bayesian inference and estimating a joint probability distribution over the variables for each time-frame. The KF is a widely applied concept in time series analysis[83].

In KF, the system equations are linear. However, if the system is not linear, then we cannot use the KF. Therefore, an adaptation of the KF is used for dealing with nonlinear systems. In this work, we consider another nonlinear state estimation approach known as the ensemble Kalman filter (EnKF). In recent times, EnKF is widely used in weather forecasting, where the models are of an extremely high order and nonlinear[84;85], the initial states are highly uncertain, and a large number of measurements are available.

4.2.3 Ensemble Kalman filter: state estimation

Ensembled Kalman Filter is introduced by Evensen (1994)[86] to overcome the disadvantages of the classical Kalman filters and Extended Kalman filters and later clarified by Burgers et al. (1998)[87]. The ensemble Kalman filter (EnKF) is a suboptimal estimator, where the error statistics are predicted by using a Monte Carlo or ensemble integration to solve the Fokker-Planck equation[88].

Like KF, in EnKF, we follow all the steps.

$$x_{k+1}^{if} = f(x_k^{ia}, \theta_k) + \nu_k^i$$

where $i = 1, 2, ...n$ and x_{k+1}^{if} is the i^{th} ensemble member forecast at time $k+1$ and x_k^{ia} is the i^{th} updated ensemble member.

With the forecasts obtained from ensembles, we can estimate the new measurements from the following equation-

$$y_{k+1}^{if} = h(x_{k+1}^{if}, \theta) + w_k^i$$

The actual observation for the $k+1^{th}$ step is perturbed with Gaussian noise to get n ensembles of the observation in the following manner-

$$y_{k+1}^i = y_{k+1} + \eta_{k+1}^i$$

where $\eta_{k+1}^i \sim \mathcal{N}(0, \Sigma_{k+1}^y)$ where $\Sigma_{k+1}^y = \frac{1}{n-1} E_{k+1}^y = [y_{k+1}^1 - \bar{y}_{k+1}, y_{k+1}^2 - \bar{y}_{k+1}, y_{k+1}^3 - \bar{y}_{k+1}.....y_{k+1}^n - \bar{y}_{k+1}]$.

Now, we form the expression for the error covariance matrix associated with the ensemble forecast. To find the covariance, we first need to find the error matrices.

The ensemble error matrix around the mean will be $E_{k+1}^x = [x_{k+1}^{1f} - \bar{x}_{k+1}, x_{k+1}^{2f} - \bar{x}_{k+1}, X_{k+1}^{3f} - \bar{x}_{k+1}.....X_{k+1}^{nf} - \bar{x}_{k+1}]$, and the ensemble output error is $E_{k+1}^y = [y_{k+1}^{1f} - \bar{y}_{k+1}, y_{k+1}^{2f} - \bar{y}_{k+1}, y_{k+1}^{3f} - \bar{y}_{k+1}.....y_{k+1}^{nf} - \bar{y}_{k+1}]$.

In the update step, we need to calculate the Kalman gain. The Kalman gain for the EnKF is expressed as -

$$K_{k+1} = \Sigma_{k+1}^{xyf} [\Sigma_{k+1}^{yyf} + \Sigma_{k+1}^y]^{-1}$$

$$\Sigma_{k+1}^{xyf} = \frac{1}{n-1} [(E_{k+1}^x)(E_{k+1}^y)]$$

$$\Sigma_{k+1}^{yyf} = \frac{1}{n-1} [(E_{k+1}^y)(E_{k+1}^y)]$$

Once we have calculated the Kalman gain, the only thing left is to update the states-

$$x_{k+1}^{ia} = x_{k+1}^{if} + K_{k+1}(y_{k+1}^{i} - y_{k+1}^{if})$$

Therefore, step by step algorithm for ensemble Kalman filter is-

Forecast step:

$$x_{k+1}^{if} = f(x_{k}^{ia}, \theta, k)$$

$$y_{k+1}^{if} = h(x_{k}^{if}, \theta)$$

Update step:

$$K_{k+1} = \Sigma_{k+1}^{xy^f}[\Sigma_{k+1}^{yy^f} + \Sigma_{k+1}^{y}]^{-1}$$

$$x_{k+1}^{ia} = x_{k+1}^{if} + K_{k+1}(y_{k+1}^{i} - y_{k+1}^{if})$$

4.2.4 Ensemble Kalman filter: dual state-parameter estimation

In dual state-parameter estimation, we need to create some parameter ensembles at the beginning. The parameter ensembles can be created depending on the empirical knowledge about parameters. We need to specify the uncertainty about parameters in the priors. Once the priors are specified, the parameter ensembles can be formed random samples from a normal distribution and some kernel smoothing[82].

$$\theta_{k+1}^{if} = \theta_{k}^{ia} + \tau_{k}^{i}$$

where $\tau_{k}^{i} \sim \mathcal{N}(0, \Sigma_{k}^{\theta})$ where θ_{k}^{ia} is the parameter ensembles calculated in the previous time step. The starting value for θ_{k}^{ia} are the ensembles of parameter priors.

If we write this in the ensemble form, it looks like-

$$\theta_{k+1}^{if} \sim \mathcal{N}(a\theta_{k}^{ia} + (1-a)\bar{\theta}_{k}^{a}, h^2 V_{k}^{a})$$

where

$$V_k^a = var(\theta^a)$$

and

$$\bar{\theta}_k^a = \frac{1}{n}\sum_{i=1}^{n}\theta_k^{ia}$$

In the above equation, a and h are some kernel smoothing parameters for the filter. Where

$$a = \frac{3\zeta - 1}{2\zeta}$$

and $\zeta = [0, 1]$ typically having a value in the range 0.95-0.99. The value of a and h has the following relation $\sqrt{1-h^2}$[82]. Forecast step: Sample θ_{k+1}^{if} from

$$\mathcal{N}(a\theta_k^{ia} + (1-a)\bar{\theta}^a, h^2 V_k^a)$$

$$x_{k+1}^{if} = f(x_k^{ia}, \theta_{k+1}^{if}) + \nu_k^i$$

$$y_{k+1}^{if} = h(x_{k+1}^{if}, \theta_{k+1}^{if}) + w_k^i$$

Update step:

$$K_{k+1}^{\theta} = \Sigma_{k+1}^{\theta y^f}[\Sigma_{k+1}^{yy^f} + \Sigma_{k+1}^{y}]^{-1}$$

$$\theta_{k+1}^{ia} = \theta_{k+1}^{if} + K_{k+1}^{\theta}(y_{k+1}^{i} - y_{k+1}^{if})$$

$$x_{k+1}^{if} = f(x_k^{ia}, \theta_{k+1}^{ia})$$

$$y_{k+1}^{if} = h(x_k^{if}, \theta_{k+1}^{ia})$$

$$K_{k+1}^{x} = \Sigma_{k+1}^{xy^f}[\Sigma_{k+1}^{yy^f} + \Sigma_{k+1}^{y}]^{-1}$$

$$x_{k+1}^{ia} = x_{k+1}^{if} + K_{k+1}(y_{k+1}^{i} - y_{k+1}^{if})$$

4.2.5 Specifying parameter priors

We need to specify the parameter priors for the filter estimation. Priors are selected based on empirical data. Prior space for the parameter is formed by sampling from the uniform distribution. From the empirical JE data, the incubation period and recovery period for humans is 5-15 days with a central tendency of 7 days[89]. Therefore, the incubation rate δ can be expressed as $1/7\ day^{-1}$ or $7/7 = 1\ week^{-1}$ as we have our data every week. The recovery period has also had a similar value. Therefore, we assume a uniform distribution for incubation rate δ and the recovery rate γ, which will give us a value close similar to the empirical value. The infection rate β can be a variable that changes widely depending on the weather factors. Therefore we choose a wide range for the uniform distribution of β. The uncertainties associated with the parameters are presented in Table 4.1.

Table 4.1: *Parameters*

Parameter	Description	Maximum	Minimum
β	Infection rate	0.01	3
δ	Incubation rate	0.8	1
γ	Recovery rate	0.8	1

4.3 Simulation results and discussion

4.3.1 Dual State-parameter Estimation and Forecast

Cumulative data of Japanese encephalitis incidence since January 2015-March 2017 is used to explore the accuracy of the dual state-parameter estimation framework for vector-borne diseases. The available data is weekly incidences of Japanese Encephalitis in Taiwan.

By combining a JE *SEIR* model with the dual state-parameter estimation framework, we can estimate parameters and states simultaneously. Figure 4.1 represents the parameter estimated from the filter with a 95% confidence interval. The parameter varies with time. Time variant parameters demonstrate a change in the model, which makes parameter estimation necessary at each step in the model and the forecast.

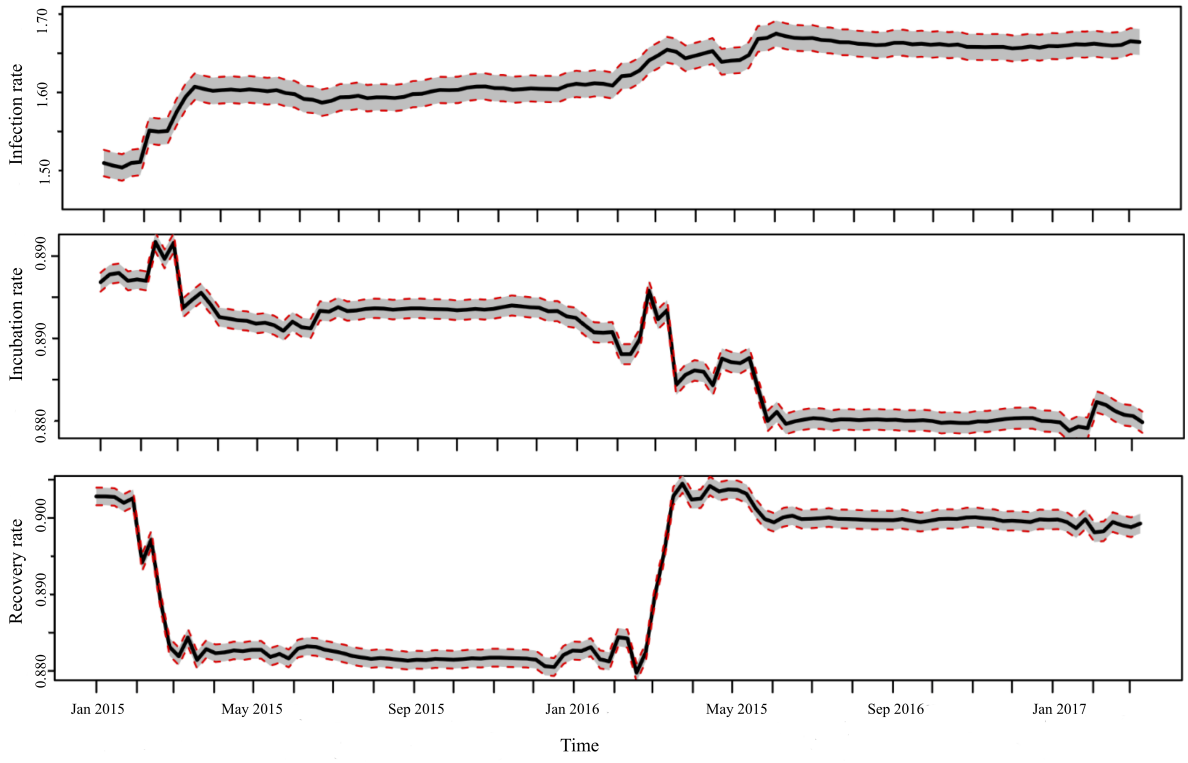


Figure 4.1: *Parameter Estimates from the dual state-parameter estimation framework with 95% confidence interval.*

Data is divided into two sections for state estimation and forecasting. We have used the first 109 weeks of data to explore estimation accuracy using EnKF. We reserved the last six weeks of data for forecasting. After 109 weeks of estimation, we use the forecasts for each new time step as the actual observation for the next step, iteratively, with the EnKF framework for estimating the parameters and forecasting. The resulting state's estimates are shown in Figure 4.2.

In Figure 4.2, we represent state estimates from the EnKF. Black dots in Figure 4.2 represents the data points used for estimating the states and parameters, while green dots represent data points reserved to check forecast accuracy.

The red line shows the estimations from the filtering framework. By visually inspecting the plot, we can say that the estimation from EnKF closely follows the actual data. Looking at the forecast, we can see that for the first four green points (four weeks), the forecasts from the filter match very closely with the actual data. However, for the last two data points, the

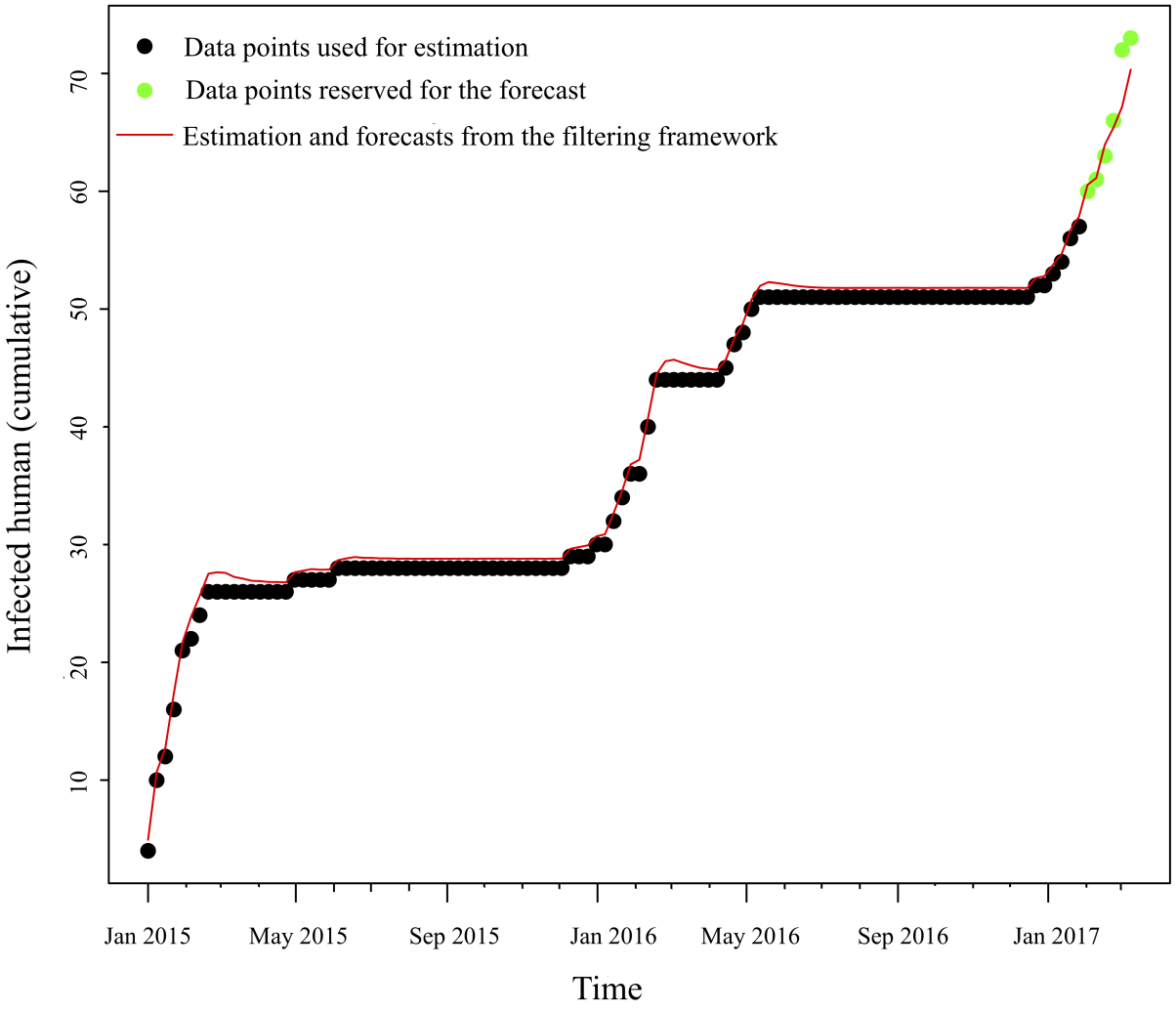


Figure 4.2: *State estimates and forecasts from dual state-parameter EnKF framework.*

filter underestimates the JE incidence estimating less infected individuals then the actual incidences. The accuracy dual state-parameter estimation is demonstrated in Figure 4.3. Figure 4.3 represents the mean squared error between the estimation and forecasts from the filter with the actual incidence data.

In Figure 4.3, black dots represent the mean squared error in estimations while the green dots represents forecast errors. Forecast errors for the first four data points have a minimal value (green points), representing greater forecast accuracy. However, the higher forecast errors for the last two weeks demonstrate the difference between the reported data and the forecasts. From Figure 4.3, we conclude that our forecasting framework is accurate for short-

to mid- (2-4 weeks) term forecasts. However, when doing long-term forecasting, the accuracy starts deteriorating with future progressions.

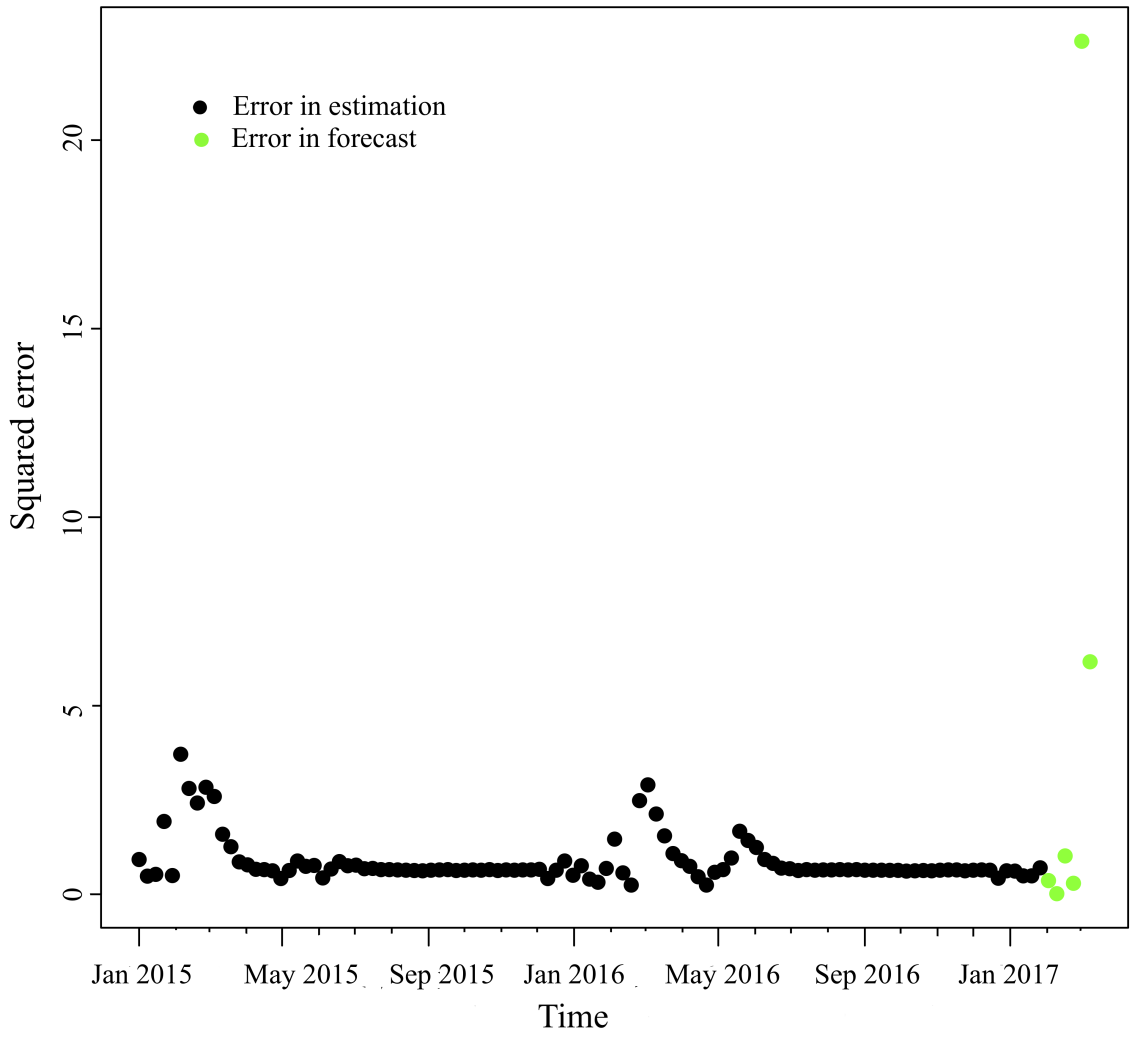


Figure 4.3: *Squared error for estimates and forecasts using EnKF framework.*

4.3.2 Application of Control Measures

The infection rate β is the only parameter that is dependent on external factors such as mosquito abundance for JE. Mosquito abundance is dependent on the weather, which makes the infection rate β weather dependent.

Therefore, if we decrease the mosquito abundance by applying some control measures such as spraying insecticide or killing larvae, this impacts virus transmission. To demonstrate the

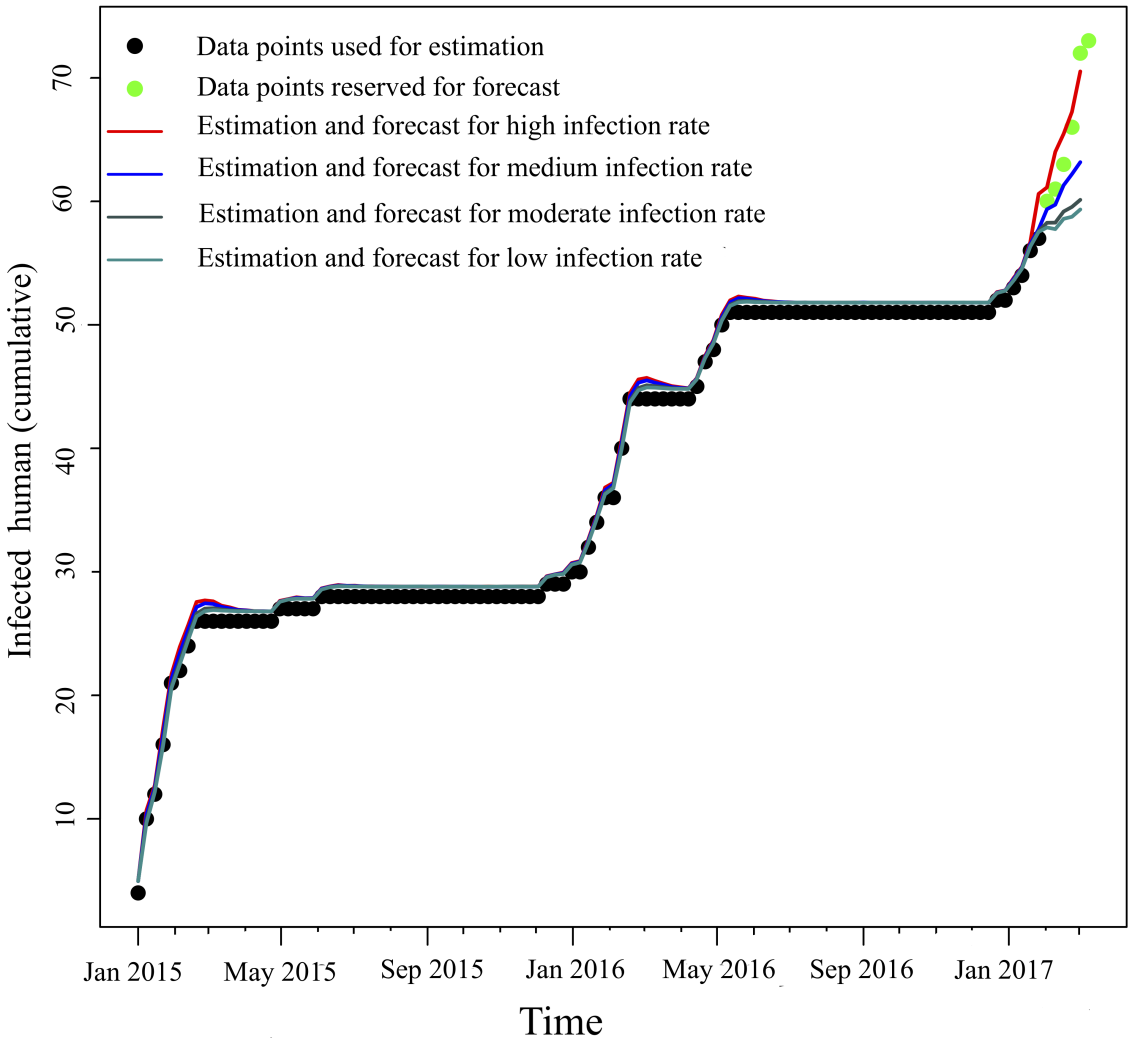


Figure 4.4: *Effect of control measures in the disease spread.*

effect of the control measure, we choose four different ranges of β, and simulation results from our filtering framework are shown in Figure 4.4. From the 4.4, we can see a significant reduction in the number of infected people for lower values of β.

4.4 Conclusions

This chapter presents the application of the EnKF for the simultaneous estimation of parameters and states for vector-borne diseases. The effectiveness of the EnKF in dual state-parameter estimation is demonstrated using Taiwan Japanese Encephalitis data. Estimated

parameters from the filter show temporal variances, which validates the need for a framework that will simultaneously estimate the parameters while doing *SEIR* state estimations. The temporal variability of the parameters is within a small range; however, that range of the parameter needs to identify for the fruitful forecast of the diseases in general. Otherwise, we will end up having an erroneous forecast. We explored the forecast horizon for JE using the EnKF. The framework is capable of accurate short- to mid-term (two to four steps) forecasts. However, for long-term forecasts, the accuracy starts deteriorating with time steps greater than four. Therefore, from our work, we conclude that our filtering framework is capable of accurate short to mid-term forecasts, even in the presence of a time-variant model. Applying control measures such as insecticides to reduce mosquito abundance (infection rate) has a conducive effect on reducing the epidemic size.

Chapter 5

Risk assessment of Ebola virus disease spreading using a two-layer temporal network[1]

5.1 Introduction

Network models for infectious disease spreading and risk assessments have opened a new era in management and containment of epidemics[91–93]. A large number of connectivity-driven network models have been proposed for various infectious diseases. These models are particularly well-suited for capturing essential system features where connections among nodes in the network are long-lived[94]. An underlying assumption with these networks is long-lived contact among individuals, which validates permanent links in the network without oversimplification[95].

Many models (both network and non-network) for infectious disease spreading have been formulated assuming homogeneous long-lived connections among individuals. Compartmental models have been used in risk assessment, estimating the basic reproductive ratio, and suggesting mitigation measures by fitting transmission dynamics with incidence data[96–103].

[1] This chapter is a reformatted and slightly modified version of our published article[90]

However, long-lived connections are not suitable for systems with rapidly changing links[95;104]. Highly contagious/infectious diseases are mostly transmitted from infected to susceptible individuals via physical contact[105]. Therefore, contagious-disease-spreading models are crucially dependent on contact structure among individuals in the network[106]. Considering a network model of the human population, the assumption of constant contact with each other is an oversimplification of reality. In general, contact structure among individuals changes with time. These changes in contacts are not entirely random; instead, there is always a pattern. For example, a contact pattern will change with changing frequency of occasional or permanent partners for sexually transmitted diseases. Therefore, the change of permanent partners can be a long-term process, and connectivity-driven network models work fine. However, when considering the occasional partner change, the connectivity-driven network will fail to capture the frequent change of partners. Several approaches have been proposed for adapting network models with changing contact patterns. One of the earliest concepts is the use of switching networks[10;107]. In this model, the contact network switches among some predetermined network structures. This model accounts for a changing contact structure with time. However, some predefined structures are required for this network model.

In the real world, contact structures are highly dynamical and evolve in time[108]. To capture dynamic contact patterns in the network, the activity-driven network (ADN) has been proposed[94]. ADNs are very powerful for studying the epidemic process when the disease dynamic and contact evolution share a common time-scale[109;110]. The activity-driven network provides the opportunity to incorporate different real-life scenarios, such as human behavior, movement pattern in the network model[110;111]. ADNs also provide means to model nodes that are likely to have contacts with the rest of the network[94;112]. Flexibility to incorporate different features makes ADN suitable for real-life network models[112;113]. The activity-driven network has been used to capture transmission dynamics of infectious diseases in *susceptible-infected-susceptible (SIS)* and *susceptible-infected-recovered (SIR)* model[108;110].

Activity-driven networks (ADNs) have been widely used for EVD spreading as well. ADNs overcome the simplifying assumption of long-lived and homogeneous contacts among individuals[93;109]. Rizzo et al. used an ADN to emulate the dynamics of EVD in Liberia and

offer a one-year prediction[93]. The effect on contact tracing on the spreading dynamics has also been quantified using an ADN[114;115]. An activity-driven network has a limitation as it randomly creates new links every time. Therefore, permanent links in the network are not considered in the activity-driven network. Several modeling frameworks have been proposed to overcome this problem with the ADN by coupling time-varying and static network components. Lie et al. proposed a static-activity coupling network[116], and Nadine et al. proposed a framework superimposing an ADN to a static backbone network[117]. These networks integrate persistent contacts with time-varying connections[116;117]. Vajdi et al. proposed a two-layer temporal network incorporating both static/permanent links and temporal/occasional links in two different layers[25].

In this chapter, we develop a novel risk assessment framework using a two-layer network with both static/permanent and temporal/occasional links in different layers. Our proposed *susceptible active-susceptible inactive-infected active-infected inactive-recovered* ($S\ S_a\ I\ I_a$R) model, and the Gillespie algorithm. The two-layer network has a permanent layer reflecting permanent contacts and a temporal layer that incorporates potential contacts. We adapted the Gillespie algorithm with the $S\ S_a\ I\ I_a$ R compartmental model and the two-layer network to see the evolution of disease spread and risk assessment. As an example of the method, we proposed a network model for Ebola virus disease (EVD) transmission in Uganda, including 23 districts based on human movement from a focal-bordering Ugandan district to Kampala. Due to a recent Ebola outbreak in the neighboring Democratic Republic of Congo (DRC), Uganda is at risk of Ebola introduction due to the entrance of an infected person. In the proposed network, the permanent layer expresses contacts among family members while intra- and inter-district contacts reflect potential contacts due to human movement. Simulation results suggest that making people aware of reduced physical contact while traveling and taking other preventive measures will reduce the number of EVD-infected humans. Results show that some districts are more vulnerable to the risk of EVD spreading than others, suggesting important guidelines for public health personnel in applying interventions and prioritizing resource allocations. Assessed risks are probabilities of infection spreading for our specific scenario based on generic and incomplete movement data, and any change in the

network will result in different risks. Therefore, risk assessments in this work are just some examples of the proposed novel risk assessment method. This chapter's main contribution is the novel, two-layer temporal network-based simulation tool for risk assessment of EVD spreading, which can be used for practical purposes when incorporating accurate movement data and model parameters.

The rest of the chapter is organized as follows. The risk assessment method section describes the two-layer temporal network, SS_aII_aR epidemics on the two-layer temporal network, and adaptation of the Gillespie algorithm for risk assessment. Application of risk assessment for Uganda EVD spreading showed an example for two-layer temporal network use in Uganda, simulation results, and discussion. We summarize our conclusions and suggestions drawn from these simulation results in the Conclusions section.

5.2 Risk assessment method

In this section, we propose a novel method for risk assessment using a two-layer temporal network, S S_a I I_a R spreading model, and the adaptation of the Gillespie algorithm for a temporal network.

5.2.1 Two-layer temporal network

We consider a two-layered network with a population of N individuals. In the two-layered model, individuals can have links among them in both layers. The intersection of the two network layers is assumed empty. We denote layers in the network as L_1 and L_2. In the first layer, L_1, links among individuals are considered permanent. Links in the second layer, L_2, are considered as potential links. In the subsequent section of this work, we call L_1 a permanent layer, while L_2 is called a temporal layer[25]. Therefore, links in the two layers will be referred to as permanent and potential/temporal links. Links in both layers are established based on some certain probability distributions. Permanent links are always active in the network, while potential links are activated with a probability only when

individuals at both ends of the link are active simultaneously. Activation of individuals is driven by an activity-firing rate (σ), as discussed in the ADN. In stochastic realization of the network, once both nodes are active, the link between them becomes active with a Bernoulli distribution having the probability P_0. The generalized structure of the temporal network is presented in Figure 5.1.

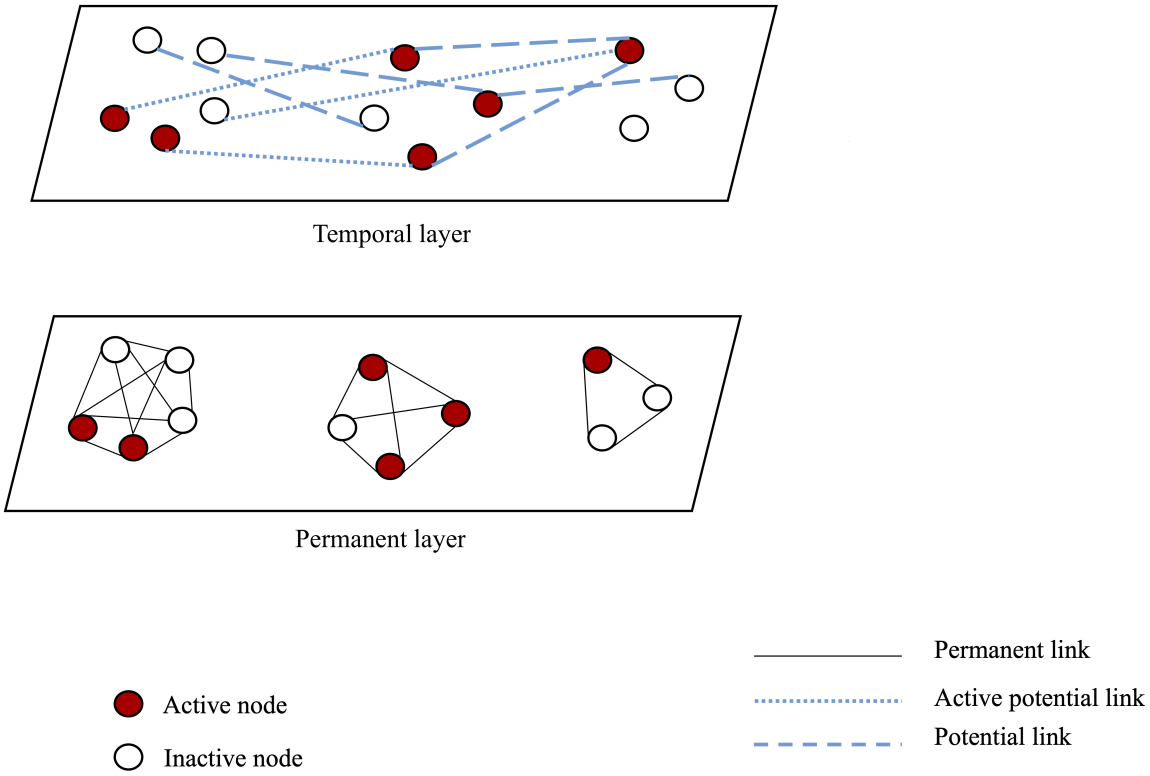


Figure 5.1: *A generalized representation of the temporal network model at a specific time t —white circles represent inactive nodes while red circles represent active nodes. Separate rectangles represent two different layers. Dark solid lines show permanent links while dashed lines show links in the potential layer. A potential link becomes active following a Bernoulli distribution with the probability P_0 when both ends of the link are active nodes.*

In the permanent layer, a link can always transmit infection, whether it is active or not. These permanent links are always present in the network. A link in the potential layer can transmit the infection only when it connects two active individuals. If there is a link between two nodes in the temporal layer, that link might be active or not, depending on the status of the nodes. The potential link becomes active with a probability P_0 only when both nodes are active simultaneously. When one node in the active link becomes deactivated, the link

vanishes. Therefore, the probability of infection transmission through the link becomes zero again. The process of a node becoming active or inactive is assumed as a Poisson process with a rate of σ. A node will stay active for an exponentially distributed period with an expected value of σ^{-1}. Thus we can assume a high value of σ for a specific node will reflect the reduced duration of an active potential link. The parameter σ is referred to as the activity rate in the subsequent sections of this chapter. Increased σ means an increased frequency of a node changing its status, i.e., becoming active/inactive. For example, we assume an individual becomes active once he starts a movement/trip and stays active until the trip is finished. Therefore, decreasing σ means increasing the trip length and a decreasing frequency of this kind of trip. Moreover, if a node does not participate in the occasional contacts, it never becomes active, and σ is set to equal zero for that node. The temporal link disappears when either of nodes in the link deactivates[25]. This temporal network is different from the widely used, activity-driven network in the contact structure among individuals. In contrast with the activity-driven network, there are permanent links, along with different temporal links, in the proposed two-layer temporal network.

5.2.2 Epidemics on two-layer temporal network

In this section, we describe the modification of the *SIR* model for our two-layer temporal network. The SIR model is a popular approach for studying infection spreading where infected people die, or eventually recover and gain life-long immunity. For diseases such as chickenpox, and EVD, the *SIR* model can describe the disease's dynamics and spread infection. In the *SIR* model, each individual is either susceptible, infected, or removed/recovered. We assume that infection and recovery processes are independent Poisson processes. A susceptible node catches the EVD infection from an infected person, and this transition happens with an infection rate β. Once a person becomes infected, he/she stays infected for a certain period, namely an infectious period. After the infectious period, individuals recover or are removed from the infection. The rate at which an infected person leaves the infected state is called the recovery rate, which is the inverse of the infection duration. The re-

covery/removal rate is denoted as γ, which has a unit $time^{-1}$. Up to this point, we have discussed the basic *SIR* model. However, for our two-layer temporal network, we must also consider the active/inactive status of the individual. Combining our temporal network model and *SIR* spreading process, we have a total of six states an individual can occupy. However, the recovered/removed population does not participate in disease transmission. Therefore, their status does not have any impact on the disease dynamics, and we can combine active and inactive recovered/removed compartments. Therefore, the model can be expressed as a five-compartment S S_a I I_aR model, where compartments are inactive susceptible, active susceptible, inactive infected, inactive infected, and recovered.

If the probability of node i occupying inactive susceptible, active susceptible, inactive infected, active infected, and recovered in the stochastic spreading process is expressed as S, S_a, I, I_a, and R, respectively, then equations for the time evolution can be expressed as follows:

$$S^{i'} = -\sigma_1^i S^i + \sigma_2^i S_a^i - \beta \sum_j a_1^{ij} S^i (I_i^j + I_a^j) \tag{5.1}$$

$$S_a^{i'} = \sigma_1^i S^i - \sigma_2^i S_a^i - \beta \sum_j a_1^{ij} S_a^i (I^j + I_a^j) - \beta \sum_j X_0^{ij} a_2^{ij} S_a^i I_a^j \tag{5.2}$$

$$I^{i'} = -\sigma_1^i I^i + \sigma_2^i I_a^i + \beta \sum_j a_1^{ij} S_a^i (I^j + I_a^j) - \gamma I^i \tag{5.3}$$

$$I_a^{i'} = \sigma_1^i I^i - \sigma_2^i I_a^i + \beta \sum_j a_1^{ij} S_a^i (I^j + I_a^j) + \beta \sum_j X_0^{ij} a_2^{ij} S_a^i I_a^j - \gamma I_a^i \tag{5.4}$$

$$R^{i'} = \gamma I_a^i + \gamma I^i \tag{5.5}$$

a_k^{ij} is the element of adjacency matrix A_k, where k=1 for the permanent layer and k=2 for the potential layer. Equations 5.1-5.5 express stochastic equations for S S_a I I_aR spreading. The variable X_0^{ij} is a Bernoulli random variable that has a value of one with a probability of P_0. This random variable is drawn each time a pair of active nodes i, j with a potential link between them occurs, regardless of their disease status. It can be translated to a real-

life expression as follows: when both nodes in a potential link are active, the link can be active with a probability of P_0. We have presented an elaborate diagram illustrating how the parameters in the exact equation drive the state transitions and state occupancy probability in Figure 5.2.

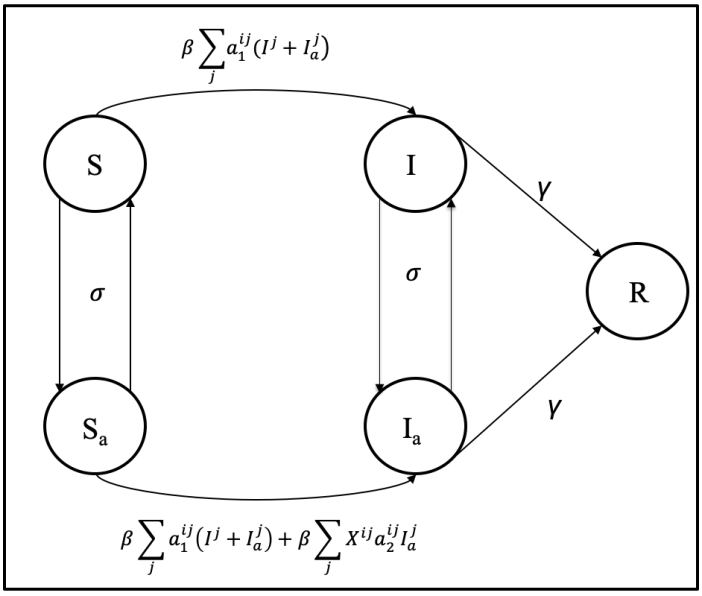


Figure 5.2: *Node-transition diagram for exact/stochastic spreading process used in this work — each circle represents a compartment whose name is written inside the circle. Expressions written over directional lines show probabilities of transition from origin to destination circle (compartment).*

Therefore, the model used in this work is stochastic, and the Gillespie algorithm was adapted accordingly for this exact/stochastic process.

5.2.3 Adaptation of the Gillespie algorithm

Gillespie algorithm has been widely used to simulate stochastic processes for static network (permanent contacts)[7;36;61;118;119] and dynamic/time-varying networks (temporal/potential contacts)[25;120]. However, our two-layer temporal network has both static and temporal

contacts. Therefore, the Gillespie algorithm was adapted to the changed network state at every time step retaining permanent contacts.

- Initialize the number nodes in the network (N), state transition matrices, the maximum number of events, and the final simulation time. Find the transition probability R_i for each node at the next time step and $R_{tot} = \sum_j R_j$, where $j = 1, 2, ...N$ and keep track of the status of the nodes (active/inactive)

- Find the time for the next event, which is exponentially distributed with a rate R_{tot}. The second step is to select a node according to the probability distribution $Pr(i) = R_i/R_{tot}$, which will make a state transition. Later we select a state where the transition will happen.

- Increase the time by time calculated in Step 2. The main difference lies in the update of transition rates for each node in the network once a transition occurs for our adaptation of the Gillespie algorithm. Nodes can change their status (active/inactive) and change their states (susceptible, infected, recovered). Therefore, every time a transition occurs, the status of the nodes needs to updated and recorded. When a node is in the permanent layer, it's transition probability has the impact of both layer's neighbors. However, when the node is in the temporal layer, it only has its neighbors' impact on the temporal layer. Therefore, the algorithm is modified to account for the network with two different layer's impact on the transition probabilities.

- Go back to Step 2 unless the stop condition is reached (maximum number of events, the final time for simulation, or $R_{tot} < Tolerance$).

5.2.4 Calculation of risk

As the spreading process in our two-layer temporal network was highly stochastic, we performed two hundred simulations for each combination of parameters. We kept track of each node's status and counted the numbers of simulations in which a particular node was in-

fected. This count was later used to calculate the risk of EVD spreading in each district. The formula to calculate spreading risk to a specific district is presented in equation (6).

$$Risk_j = \frac{\sum_{n=1}^{N_j} I_n}{N_j N_{simulation}} \tag{5.6}$$

Where, $Risk_j$ = Ebola spreading risk of district j, I_n = Number of simulations where node n is infected, $N_{simulation}$= Number of simulations, N_j= Total number of population in j^{th} district, and $j = 1, 2, 3, ..., 23$.

Once we calculate the risk, we can use any mapping software such as ArcGIS software to create risk maps. Risk maps provide a visual representation of spreading risks.

5.2.5 Calculation of confidence interval

Our simulation framework is an event-based algorithm that randomly chooses the time when the next event will occur and what that event will be. Therefore, these events are not homogeneously time slotted. We perform a time regularization, where the temporal window for simulation is divided into equal periods, and the number of events is calculated within each period. We obtained the size/number of individuals in each compartment from events in each period. We performed 200 simulations for each parameter set and recorded the number of individuals in each compartment for all periods. Later, we found the range within which 95% of the simulation results fall in each period.

5.3 Application of risk assessment for Uganda EVD spreading

We applied our risk assessment method for EVD in Uganda using a generalized movement pattern and some specific model parameters. First, we formulated the two-layer temporal network for Uganda. Later we used the network with the SS_aII_aR model and the modified

Gillespie algorithm to track the number of infected humans after a certain period and assess the risk of EVD spreading to different spatial locations. Simulation results and discussion for specific scenarios are presented later in this section.

5.3.1 Two-layer temporal network for Uganda

A recent Ebola outbreak in the DRC created a possibility of EVD infection in Uganda. We proposed a two-layer temporal network for EVD spreading in Uganda. We have observed the focus of Ebola preparedness by the ministry of health and partners to select a possible point of entry for an Ebola patient from the DRC. We found the Kasese district was at high risk of an EVD- infected person's entry point. Therefore, we chose this district as our point of entry. Once an infected person from the DRC entered into Uganda through the bordering districts, he or she met susceptible people in that location. EVD-infected persons, being highly contagious, can spread the infection to people they meet. People move from one location to another for different reasons, and there is always a pattern for this movement. We created a network based on people's movement for different purposes such as fish trade, cattle trade, and general movement. The movement pattern and districts in the movement paths were obtained from confidential data provided by the ministry of health in Uganda. We used these data to formulate an example of a human movement network for some selected districts in Uganda. Human movement from one location to another is largely motivated by three different purposes in Uganda:

1. Fish traders move in a southward direction from the point of entry.

2. General movement for shopping, visiting relatives, searching for work, or traveling for various purposes starts at the point of entry. This movement goes all the way to the capital city of Kampala. People mostly travel from rural areas to neighboring big cities. They meet other people there, and this results in a certain mixing among individuals from different locations. This mixing happens throughout this movement path. Several of these movement paths are presented in Figure 5.3 with green arrows.

3. Limited movements due to cattle trade are mostly local or between neighboring districts. Long-distance cattle trade happens at the commercial scale and does not include much movement of people as they move mostly via an organized transportation system.

We created a specific network for human movement from Kasese to Kampala based on general movement information. This network was used to show an example of the method developed for risk assessment. As we described in our temporal network section, we assumed two layers in the network. The permanent layer consisted of contacts of each individual within each household. Each family in Uganda has an average of six children[121]. Including parents, we roughly assumed around eight people within each household. Therefore, we assumed a population distribution where each household had five to ten members.

Within each household, we assumed permanent links among family members. Therefore, links within each household constructed the permanent layer in the Uganda EVD spreading network. The potential layer was formed by incorporating previously discussed human movement. An individual becomes active once he or she is in movement and stays active until he or she finishes the movement. Once a node is in movement (active), its potential link can be activated if the other node in the link is active. This can be explained as follows: node i has potential links with a set of nodes named J throughout the whole network. Therefore, once node i is active, potential links with any of the active nodes in J can be activated following a certain probability. This link-activation structure is crucial in proper representation of the contact structure. Two moving nodes (active) can meet each other in different places such as transportation, marketplace, visiting sites. Usually, the movement pattern between individuals follows a general structure. Within each location, a probability exists that individuals encounter each other for various purposes. Inter-location contacts also follow some structure rather than being completely randomized. People usually flock into big cities or towns nearby where they encounter local active people or others coming to that location. When this happens, if two of these active individuals have a potential link, that link is activated with a certain probability distribution (Bernoulli distribution with probability P_0).

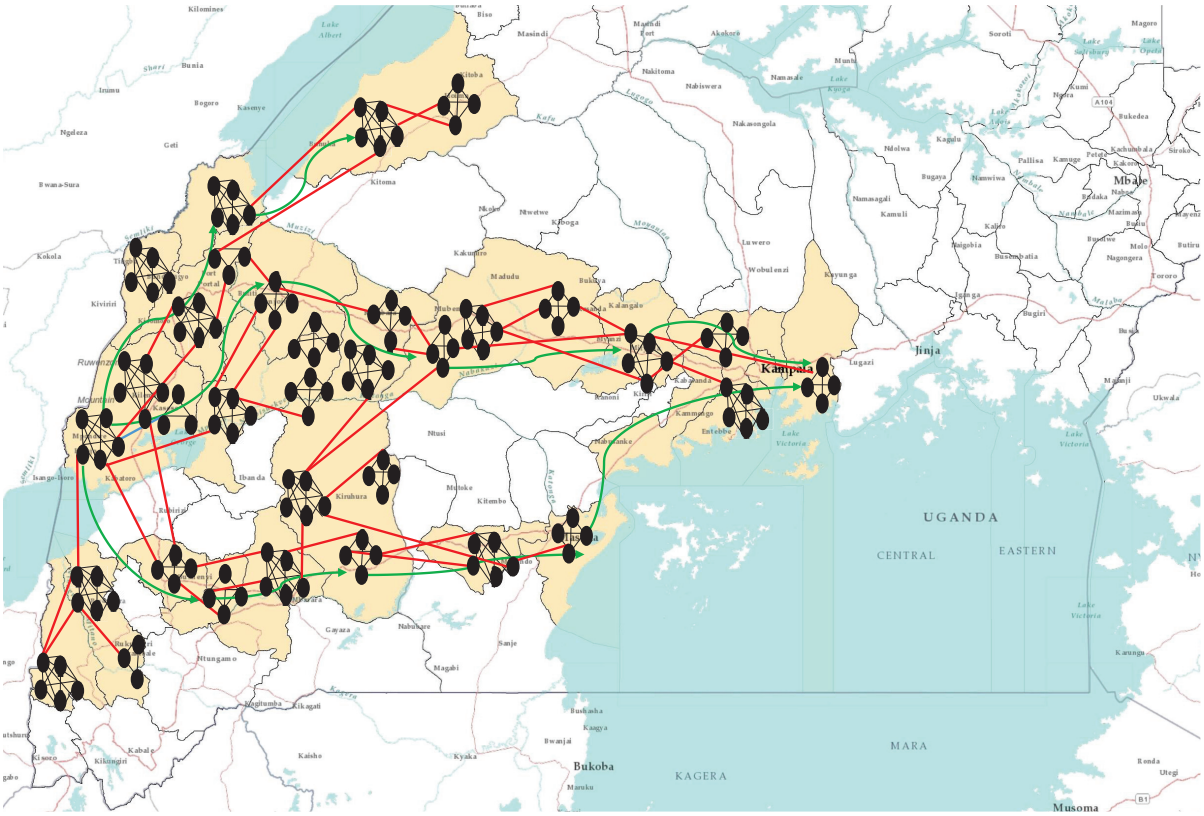


Figure 5.3: *Two-layer temporal network for EVD spreading in Uganda. The districts considered for our network model are colored in baize (23 districts). Small black oval shapes represent individual human beings. This figure does not represent the actual network; rather, it is a visual representation of the two-layer network. Each cluster of a human represents a household, and black lines among them represent permanents contacts. Red lines represent contacts in the potential layer. These links only become active when both individuals in the link are active simultaneously. A link represents the possibility of pathogen transmission. The green directional arrow represents the directions of human movements.*

A zoomed up view of a portion of the network (Kasese district) is presented in Figure 5.4. This shows a scaled-down version of the actual network for visualization purposes.

Districts used in our Uganda network are presented in Table 5.1. We used the centroid of each district to find distances between districts while formulating the network. However, this is one realization of the network built using available movement data. Some Ugandan districts are not included in our assessment; therefore, there may have been districts at high risk that were excluded from the risk assessment performed in this study. We scaled the population of each district in our network by 1,000 for computational purposes. As our main purpose was to evaluate the risk of EVD, this scaling greatly reduced the computational

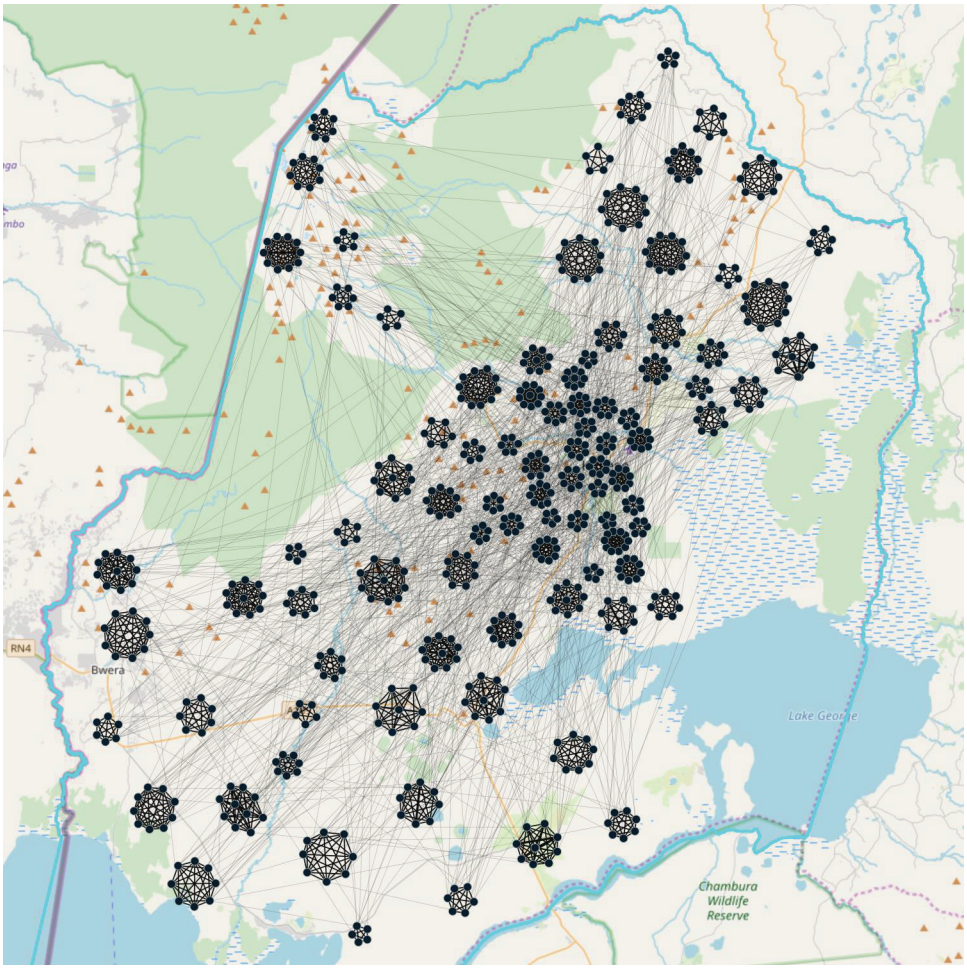


Figure 5.4: *Zoomed view of a district in EVD spreading network in Uganda. We chose the Kasese district for this visualization. However, the population in the specific district is scaled down for a comprehensible representation of the network. Small clusters represent individual households (permanent layer), and lines between clusters represent temporal links. We created the network using Gephi-0.9.2 (https://gephi.org/) and the map using ArcMap 10 (http://desktop.arcgis.com/en/arcmap/). We have used Inkscape 0.92.3 (https://inkscape.org/) to superimpose the map and network together.*

complexity. We have incorporated previously discussed movement patterns along different paths from one point of entry to Kampala in the network. These movements and their directions are also shown in Figure 5.3. We considered all districts that were in movement paths of any nature toward Kampala from Kasese. Although all bordering districts were at risk of Ebola introduction, we focused on demonstrating the application of our method when the initial infections were in the Kasese district.

In summary, our network for Ebola spreading in Uganda consisted of 23 districts where

Table 5.1: *Districts considered in our Uganda two-layer temporal network*

Districts	Population
Bundibugyo	224387
Bunyangabu	181200
Bushenyi	234440
Hoima	572986
Kabarole	469236
Kampala	1507080
Kamwenge	414454
Kanungu	252144
Kassanda	100038
Kasese	694992
Kyegegwa	159800
Kyenjojo	422204
Lwengo	267300
Masaka	297004
Mbarara	472629
Mityana	328964
Mpigi	250548
Mubende	684337
Mukono	594804
Ntorko	70900
Rukungiri	314694
Sheema	180200
Wakiso	1997418

the permanent layer incorporated contacts among individuals within a household. The potential layer reflected contacts between individuals when both ends of the link were active during movement and had a possibility of pathogen transfer between them.

5.3.2 Simulation setup

Upon formulating the two-layer network, we performed simulations with the SS_aII_aR model for EVD transmission using the Gillespie algorithm[25;31]. We conducted our simulations with two major goals — to observe the progression of EVD spreading with time throughout the network and evaluate the risk of spatial spreading for this specific scenario. In our simulations, we had four different parameters. They were the rate at which individuals

become active and inactive, generally called activity rate σ; the infection rate β; the recovery rate γ; and the probability P_0 of an active link between two simultaneously active individuals in the potential layer. These values are hard to estimate, and due to high stochasticity in the people's movement pattern, movement parameters, i.e., (σ), cannot be expressed with a single value. The infection rate β is also variable and takes different values for different outbreaks depending on the contact pattern among individuals. We performed simulations using multiple values of each parameter to explore the sensitivity of epidemic size and transmission risk. For each parameter set, we calculated the number of infected individuals with time and created risk maps.

We presented the number of infected individuals in each time step for each parameter set. As we had multiple parameters, we chose the value of P_0=0.7 and 0.1, which represented a 70% and 10% chance of an active potential link between two active nodes. We chose σ= 0.1 and 0.5 for the activation/deactivation rate (activity rate) and $\frac{1}{\sigma}$ is the average time an individual is active. The value of σ has the following real-life explanation: a particular individual becomes active in every $\frac{1}{\sigma}$ days, and once active, he or she stays active for the next $\frac{1}{\sigma}$ days. For example, σ= 0.1 represents the frequency of a node being changing its state (inactive to active, active to inactive) every ten days.

5.3.3 Results and discussion

We initiated our simulation with a single, active infected person in the Kasese district. We tracked each node's status in the network for 150 days to see how each node was changing its status (inactive susceptible, active susceptible, inactive infected, active infected, and recovered). At the end of the simulation time (150 days), the total number of infected people in the outbreak was calculated, and the risk of a specific node being infected during the outbreak was assessed.

To measure the progression and severity of EVD spreading, we tracked the number of infected people and a cumulative number of infected people for 150 days. For any infection spreading, some important measures are the size of the peak infection (maximum number

of simultaneously infected individuals), time to reach that peak, and the total number of infected within an outbreak (epidemic size)[36]. We designed our simulation to track both the number and the cumulative number of infected at each time step. Our simulation results are presented in the subsequent parts of this section. We had chosen values of infection rate β to explore a varying range of transmission potential given contact with infected individuals. We have calculated the 95% confidence interval from 200 simulations for the number of humans in each compartment at each period and final epidemic size. For easier demonstration and comparison, we have presented simulation results without a confidence interval in the Results section.

Figure 5.5 represents the number of cumulative infected and infected humans for P_0 =0.7 and σ=0.5, and a varying range of infection rate β.

Within Figure 5.5, the top graph represents the average cumulative number of infected while the bottom plot represents the average number of simultaneously infected humans at different time steps.

From Figure 5.5 (P_0=0.7 and σ=0.5), it is evident that with the increase of β, infection size increases rapidly. An increase of β from 0.2 to 0.5 increases infection size from 2,459 to 6,634. Therefore, a very small increase in transmission causes a huge increase in the number of people infected with EVD. We assumed a 70% chance of pathogen transmission once an active infected individual comes into contact with another active susceptible person in the potential layer. Also, in this simulation set, the potential layer is assumed to be highly active/mobile. People are assumed to become active and inactive within an average period of two days.

The number of simultaneously infected people at a certain time is very important for public health personnel[122]. More infected people means increased preparation of hospital beds, doctors, and medical supplies[123]. Therefore, once an outbreak occurs, it is important to have an idea about the maximum number of simultaneously infected people (size of the peak infection)[36]. In Figure 5.5, the bottom plot shows simultaneously infected people at each time step. In the plot, peak infection size also increases with β. However, when β=0.2, the peak is not very pronounced, and the peak infection size is around 600. However, for

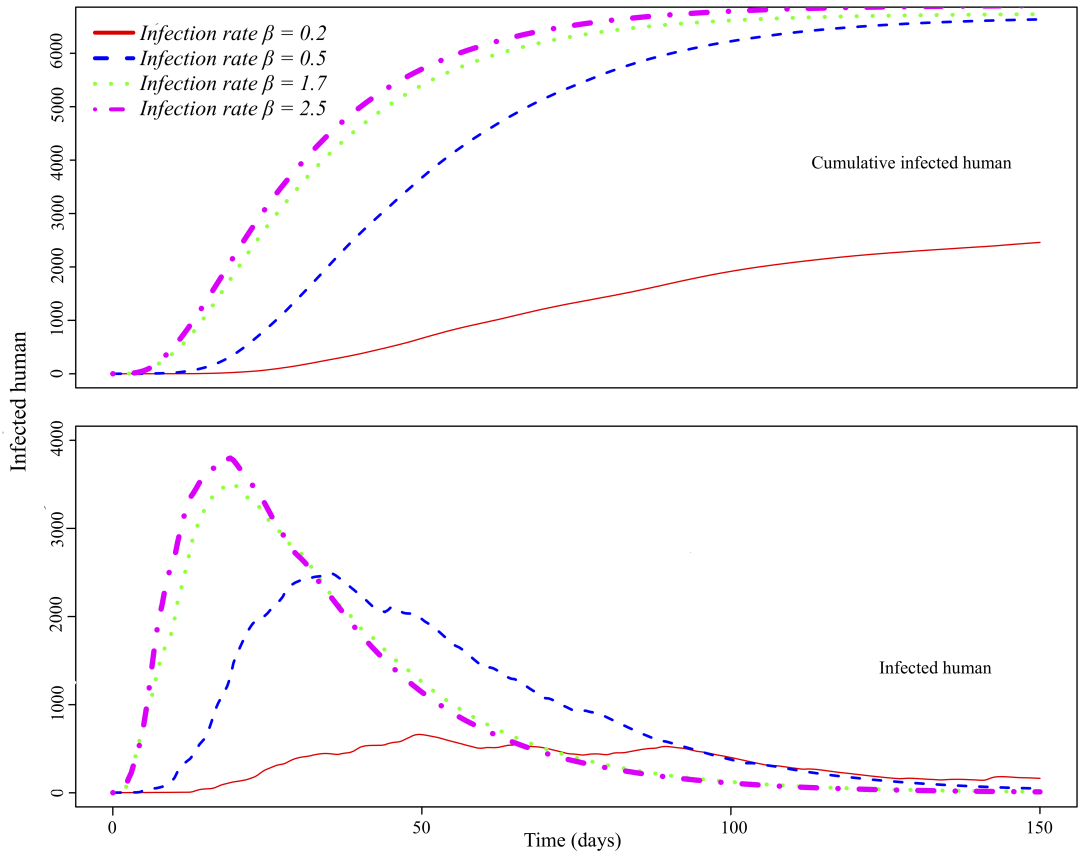


Figure 5.5: *(Top) Average number of cumulative infected and (bottom) average number of infected humans in the Uganda Ebola network for P_0=0.7, σ=0.5 —four different colors represent the average number of infected humans for four values of β. Red, blue, green, and magenta lines correspond to the number of infected humans for β = 0.2, 0.5, 1.7, and 2.5, respectively.*

other values of β, the peak is more pronounced, and the peak infection size is more than 2,500 for all other values of β. The time to reach the peak for β=0.2 is around 50 days. With the increase of β, the infection plot becomes skewed to the left, meaning faster arrival at peak infection size. Faster arrival to peak infection and greater value of peak infection size indicates a widespread epidemic outbreak. Therefore, simulations for this specific parameter set indicate a widespread and severe outbreak for β larger than 0.5, where more than 50% of our total population in the network becomes infected.

For each parameter set, we had estimated the risk of EVD spreading to different spatial locations from a single infected individual at the Kasese district. We had created risk maps using ArcGIS for EVD spread to distant locations. We had used an equation for risk ex-

plained previously to calculate the risk. The values of risk are classified into five different categories, as presented in Table 5.2.

Table 5.2: *Classification of risk for our spatial locations based on the value of risk parameter.*

Risk	Value of risk parameter
No risk or not considered	0
Low risk	$0 < \text{risk} \leq 0.2$
Moderate risk	$0.2 < \text{risk} \leq 0.4$
Medium risk	$0.4 < \text{risk} \leq 0.6$
High risk	> 0.6

Risk maps in Figure 5.6 show all neighboring districts in the southern part of Uganda, although we have considered 23 districts. Therefore, some districts on the maps had not been considered in our two-layer temporal network and, hence, have not been assessed for risk.

Figure 5.6 shows a risk map for four selected values of β and P_0= 0.7, and $\sigma = 0.5$. The map is colored with a monochromatic color gradient, which increases with increased risk. Therefore, districts that are white on the map are either not at risk (if considered in the network model) or not considered in the network model.

From the risk map for β=0.2 in Figure 5.6, we can see that all of our selected locations are at low or moderate risk as the infection rate is very low. However, some districts are at comparatively higher risk than others are. For this specific scenario, with β=0.2, Bundibugyo, Bushenyi, Kyegegwa, Kyenjojo, Masaka, Mpigi, and Sheema districts are at higher risk than other districts in our network. However, the districts mentioned above are at a moderate risk of Ebola spreading while other districts are at low risk for β=0.2. Increasing the value of β increases risks proportionately, which can be seen from Figure 5.6. Color gradients increase in risk maps with increasing β. In the risk maps for other values of β, a similarity in risk is observed. This is evident from similar color gradients of districts on all three maps for β=0.5, 1.7, and 2.5, which can be explained from the similar epidemic size and comparable infection peaks for these values of β (Figure 5.5). Therefore, we have similar infection spreading in these cases, and assessed risks are similar (not the same, but

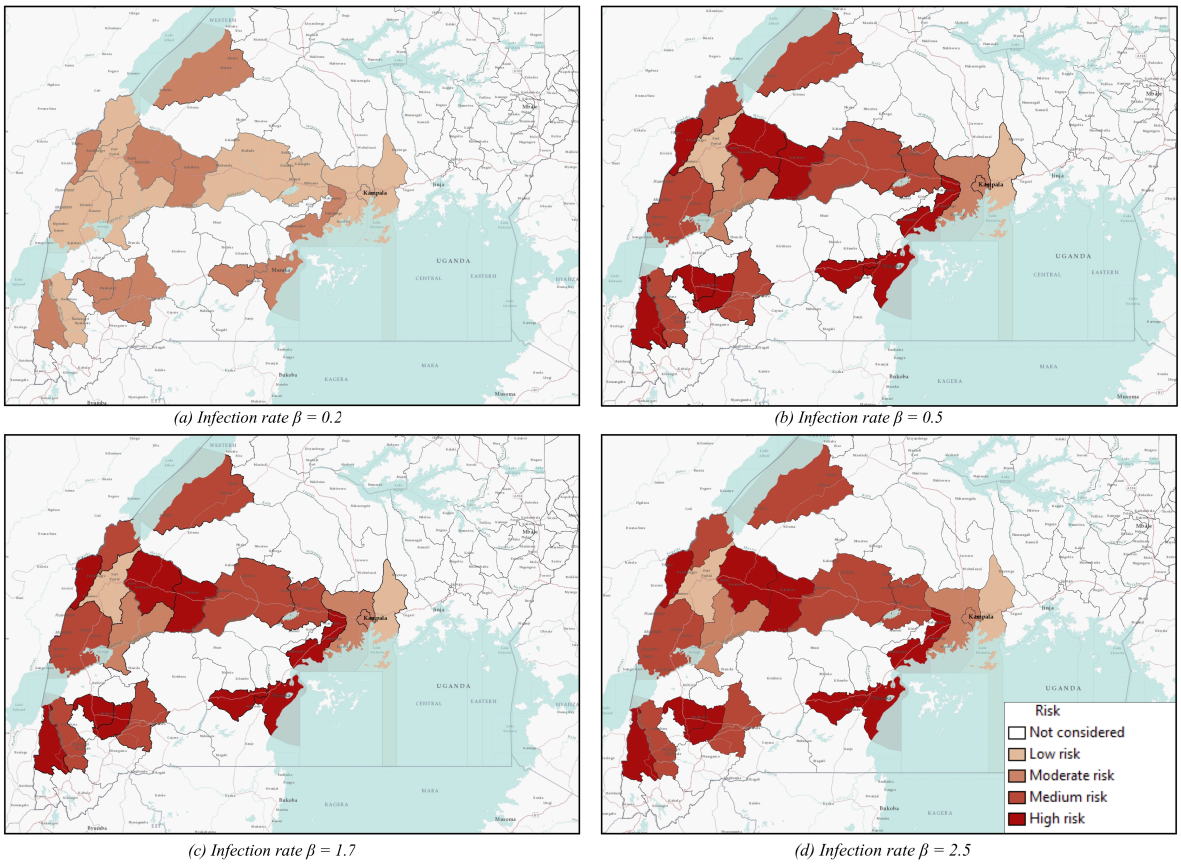


Figure 5.6: *Risk map of Ebola spreading within selected 23 districts in Uganda for P_0=0.7, σ=0.5, and (a) β=0.2, (b) β=0.5, (c) β= 1.7, (d) β = 2.5. The map is colour coded according to the risk of Ebola spreading.*

they are in the same interval as presented in Table 5.2). There was an increasing trend in the value of calculated risk parameters with an increase of β, although, on the map, they are included in the same interval. Districts in the temporal network for β= 0.5, 1.7 and 2.5 with assessed risks are presented in Table 5.3.

Table 5.3: *Districts in the network and their associated risks for EVD spreading.*

High risk	Bundibugyo, Bunyangabu, Bushenyi, Kanungu, Kyegegwa, Kyenjojo, Lwengo, Masaka, Mpigi, Sheema
Medium Risk	Hoima, Kasese, Mbarara, Mityana, Mubende, Ntoroko, Rukungiri, Kassanda
Moderate risk	Kampala, Kamwenge, Wakiso
Low risk	Kabarole, Mukono

Table 5.3 shows that 10 of 23 districts are at high risk of EVD spreading during an outbreak. With a current outbreak in the DRC, it is expected that bordering districts will be at high risk of EVD spreading. However, our simulation results can incorporate movement

data that shows non-bordering districts can also be at high risk due to the infected persons' movement. This can be easily seen from Figure 5.6, where some non-bordering districts demonstrate a high or medium risk of EVD spreading.

Figure 5.7 represents the number of infected individuals for P_0 =0.7 and σ=0.1. $\sigma = 0.1$ means individuals are likely to become active and inactive with an exponentially distributed time, which has an average of 10 days. Therefore, decreasing σ decreases the probability of a human becoming active while increasing the time of him/her staying active. Peak infection increases from 450 to 1,700 for our selected values of β=0.2 to 2.5.

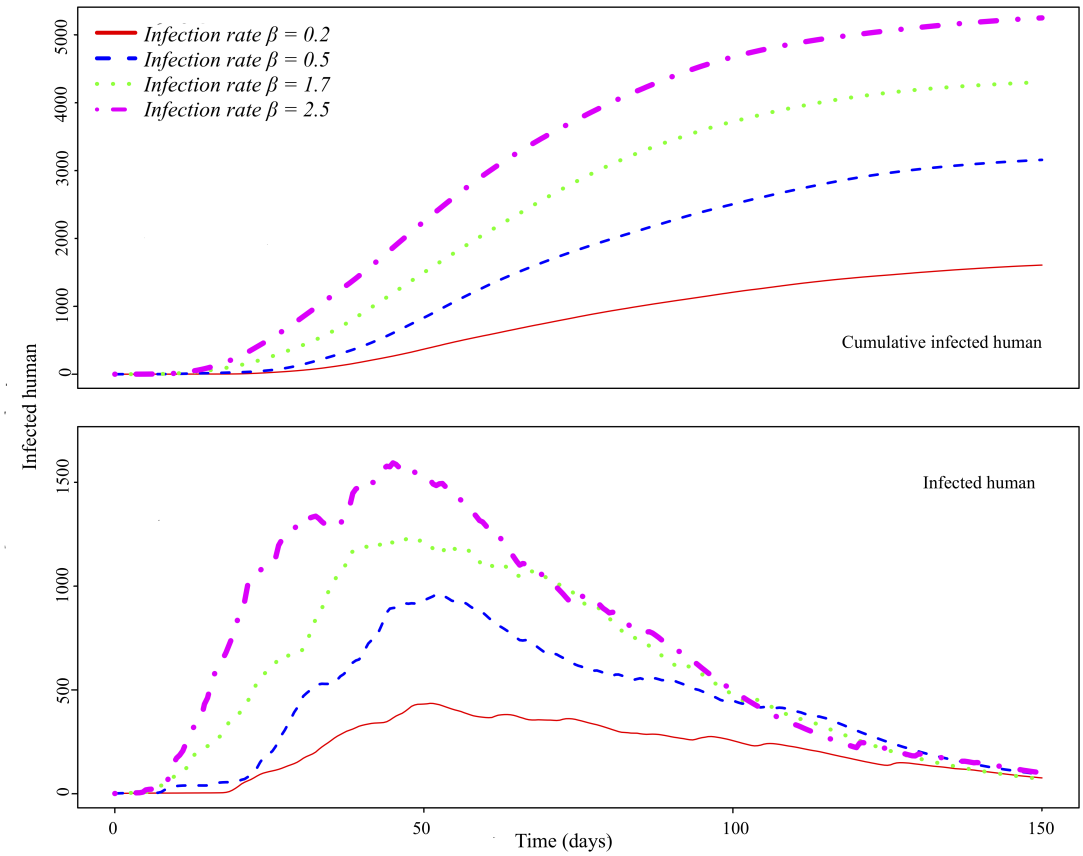
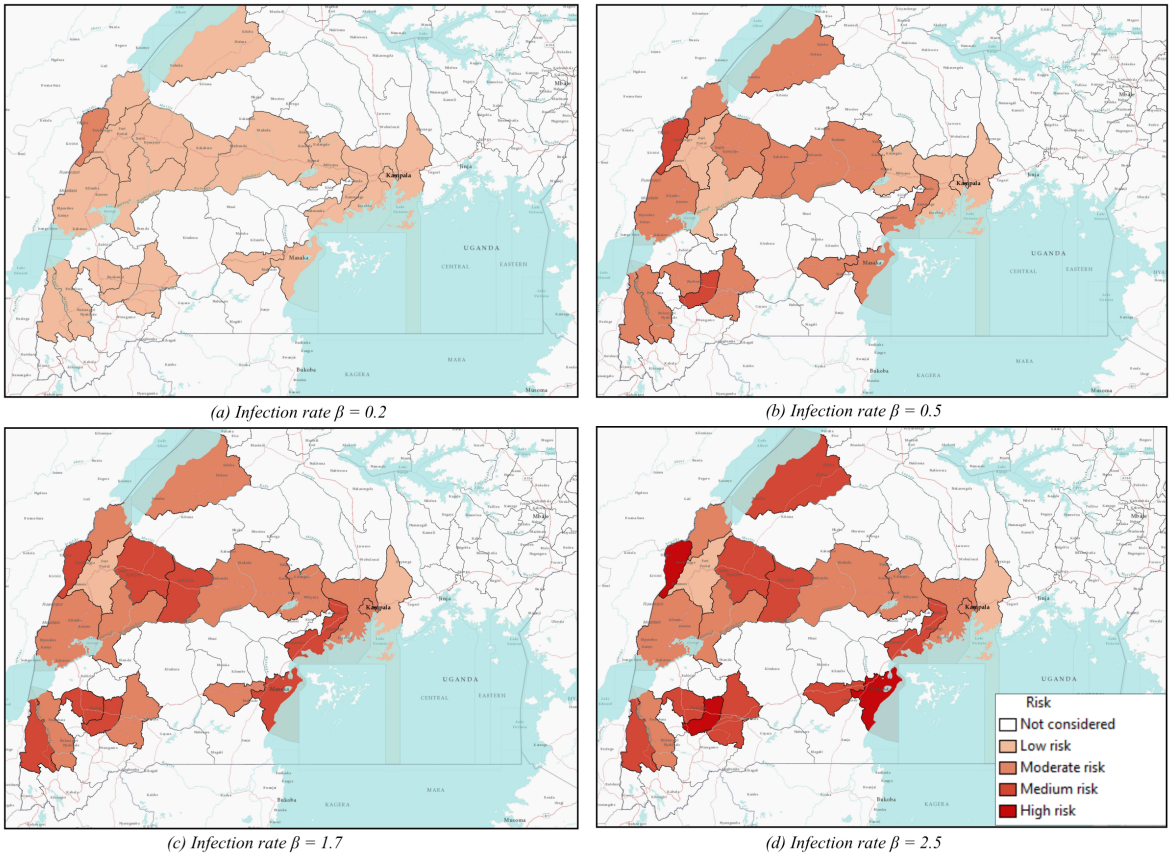


Figure 5.7: *(Top) Average number of cumulative infected and (bottom) average number of infected humans in the Uganda Ebola network for P_0=0.7, σ=0.1 — four different colors represent the average number of infected humans for four values of β. Red, blue, green, and magenta lines correspond to the number of infected humans for β = 0.2, 0.5, 1.7, and 2.5, respectively.*

Comparing simulation results presented in Figure 5.5 and 5.7 for similar values of β, epidemic size, as well as peak infection, is always higher for higher values of σ. Higher values

of σ can be translated to increased human movement in our network. As we are assuming an active person is a traveling person, a higher σ for a person means frequent short trips, while lower σ means infrequent longer trips. Higher σ decreases the time a human will be in the active (mobile) state, while a lower value indicates the longer length of an individual staying active[25]. Therefore, it is convenient to assume that while an infected individual is active for a longer period, this will eventually spread the infection to more active individuals in the potential layer[124]. However, our simulation results show otherwise. An increase in the value of σ increases infection size as well as peak infection size (Figure 5.5 and Figure 5.7). Therefore, our simulation results show the frequency at which individuals become active or inactive dominates over the individual's stay active. This is evident from the larger size of the epidemic from the higher value of σ. For β=0.2, the cumulative number of infected after 1,607 is when it is 59, and we have σ=0.5. The value of the cumulative number of infected is higher for other cases as well. The infection reaches its peak slowly (approximately 50 days) irrespective of the value of the infection rate. Therefore, the lower value of the activity rate σ means a lesser number of simultaneously infected people and a slower spread of infection within the spatial locations (Figure 5.6 and Figure 5.8). As we discussed earlier, a lower value of σ reflects the reduced movement of people. From comparisons between simulation results in Figure 5.5 and Figure 5.7, it is evident that human movement is critical in the severity and speed of EVD spreading. As frequent human movement (frequent short trips) spreads EVD very quickly, reduced human movement may minimize the severity of the EVD spread[125]. This is also evident from the risk map shown in Figure 5.8 for reduced σ. From the map, we can see that Bundibugiyo, Sheema, and Masaka are three districts most at risk of Ebola spreading for our specific network model. For these three districts, the risk of spreading is comparatively higher than other districts, even when the value of the infection rate is very low. For example, these districts are in a moderate-risk zone while others are in a low-risk zone for β=0.2. However, with the increase of β, the value increases, and for our highest value of β=2.5, these three districts are at a high risk of Ebola spreading. Table 5.3 presents districts in the temporal network for β= 2.5 with assessed risks.

During the network creation, some districts were assumed possible mixing places, which



(a) Infection rate β = 0.2 *(b) Infection rate β = 0.5*

(c) Infection rate β = 1.7 *(d) Infection rate β = 2.5*

Figure 5.8: *Risk map of Ebola spreading within selected 23 districts in Uganda for P_0=0.7, σ=0.1, and (a) β=0.2, (b) β=0.5, (c) β= 1.7, (d) β = 2.5. The map is colour coded according to the risk of Ebola spreading.*

makes them more vulnerable than others. Therefore, simulation results and evaluated risks were obtained according to the network structure. An example of dependency on the network structure is evident from Figure 5.6 and 5.8, showing the district of Kabarole at low risk for all values of β, despite being a bordering district to DRC. This low risk for Kabarole reflects the fact that this district was not considered as a mixing place in our network. This demonstrates our method's adaptability to specific data about each location in the network.

Table 5.4: *Districts in the network and their associated risks for EVD spreading.*

High risk	Bundibugyo, Masaka, Sheema
Medium Risk	Bunyangabu, Bushenyi, Hoima, Mbarara, Mityana, Mpigi, Kyegegwa, Kyenjojo, Lwengo, Kanungu
Moderate risk	Kampala, Kamwenge, Wakiso, Rukungiri, Ntoroko, Mubende, Mityana, Kasese, Kassanda
Low risk	Kabarole, Mukono

Comparing Table 5.3 and Table 5.4, it is evident that when we have a lower σ, risk

decreases significantly. For the same values of β, lower σ reduces the risk of Ebola spreading. When β=0.5 and σ=0.5, there are 11 high-risk districts. Decreasing the value of σ to 0.1 results in only three high-risk districts for β=2.5. For other values of β and σ=0.1, none of the districts in our network model are at high risk. Therefore, reducing human movement has shown a significant decrease in EVD spreading. However, reducing human movement is not practical, as it cannot be controlled. Therefore, we focus on the parameter P_0, which expresses the possibility of pathogen transfer via an active potential link. P_0 can be expressed as the probability of an actively infected person spreading the virus to an active susceptible person. The value of this parameter is dependent on direct physical contact between the infected and the susceptible. In our previous set of simulations, we used P_0=0.7. If we can decrease the possibility of physical contact among humans in case of an outbreak, it would be equivalent to a reduced possibility of virus transfer. To observe the impact of reduced physical contact, i.e., lower P_0, we conducted a simulation for a P_0=0.1 reflecting only a 10% possibility of pathogen transfer via an active link while one of the humans is infected.

Decreasing the probability of EVD spreading (i.e., P_0) to 10% from 70% via a contact in a potential layer significantly reduces infection size as well as numbers of simultaneously infected humans. However, while changing the P_0, similar values of σ are used as before.

Figure 5.9 shows cumulative infected humans for σ=0.5 and P_0=0.1. For our lowest value of β=0.2, the cumulative number of infected humans is 45 after 150 days. Increasing β increased the value of cumulative infected humans to 1,274 for β=2.5. Therefore, the number of cumulative infected humans is very low compared to cumulative infected humans for similar values of β with P_0=0.7. Besides, there is no pronounced single peak, while the maximum simultaneously infected people go to around 400 for the highest-used value of β. Therefore, decreasing P_0 reduces the number of infected humans and, thus, the severity of EVD spreading.

Figure 5.10 shows risk maps for P_0= 0.1 and σ=0.5. It is evident from the maps that all districts in our network are at low risk for EVD spreading for selected values of β.

Further decreasing the value of σ to 0.1 decreases the cumulative number of infected humans as well as peak infection size. Figure 5.11 shows that when β is 0.2 and 0.5, the

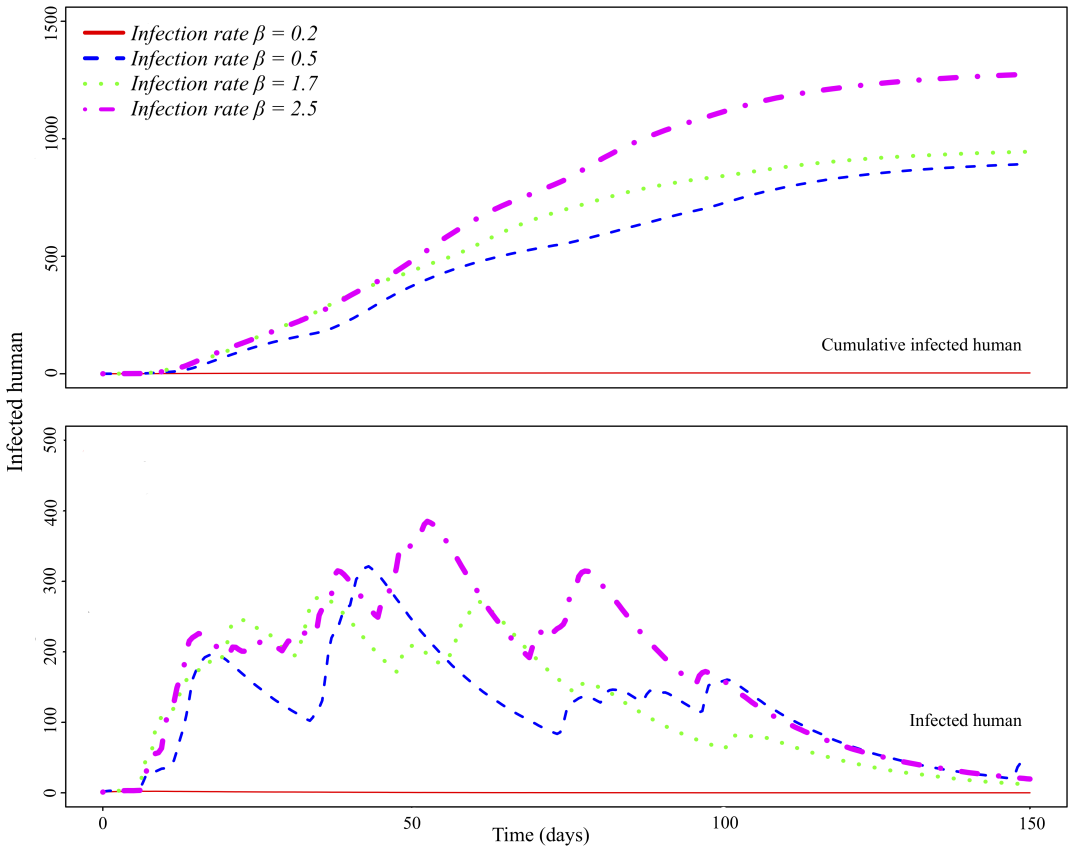


Figure 5.9: *(Top) Average number of cumulative infected and (bottom) average number of infected humans in the Uganda Ebola network for P_0=0.1, σ=0.5 —four different colors represent the average number of infected humans for four values of β. Red, blue, green, and magenta lines correspond to the number of infected humans for β = 0.2, 0.5, 1.7, and 2.5, respectively.*

EVD does not spread at all and stays within the outbreak location with only two to three infected persons. However, increasing the value of β increases the number of infected humans, but the cumulative number of infected remains less than 150 even for our highest value of infection rate ($\beta = 2.5$).

Figure 5.12 shows no risk of EVD spreading for β=0.2 and 0.5. However, with increasing β, all our districts are at a low risk of EVD spreading.

Summarizing simulation results and risk maps presented in Figures 5.6-5.9, we can see a significant reduction in the epidemic size and simultaneously infected humans when P_0=0.1. Therefore, a reduction in the probability of pathogen transfer via a potential link i.e., reduced physical contact between humans while they are active/mobile, greatly reduces the number

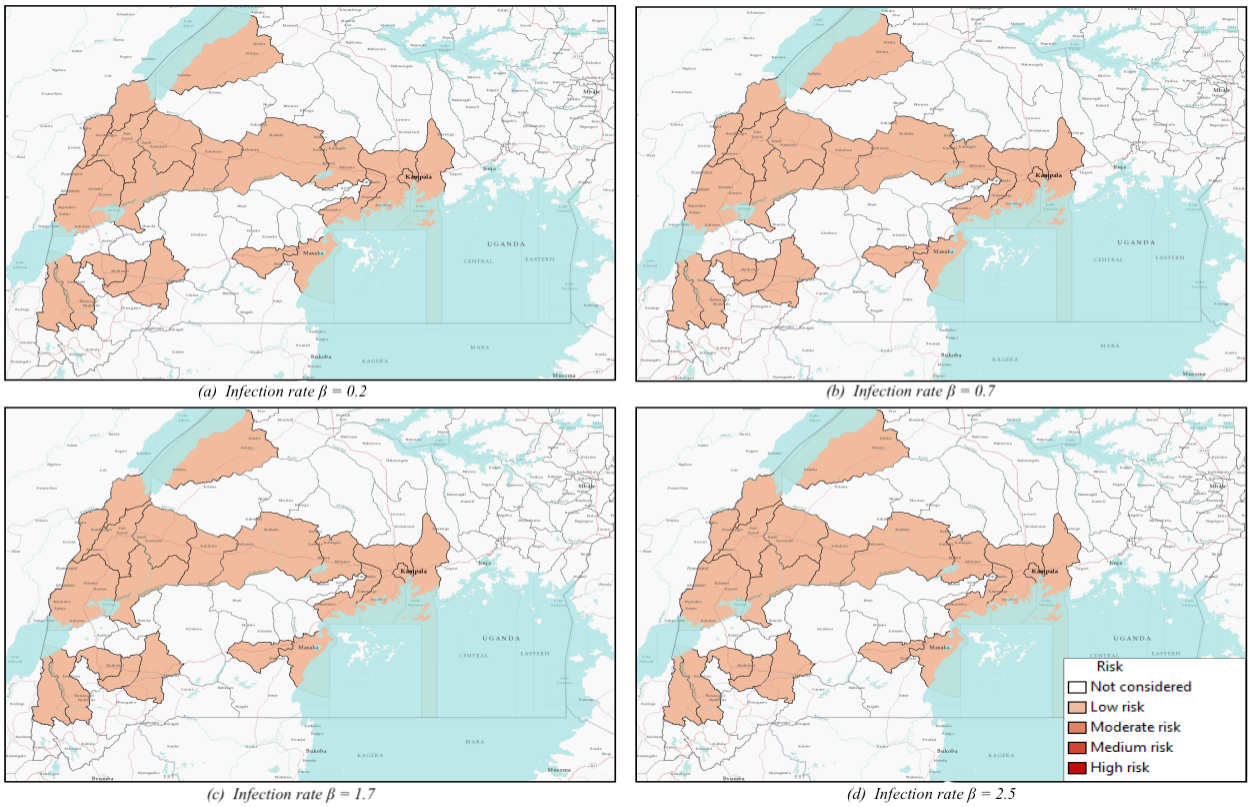


Figure 5.10: *Risk map of Ebola spreading within selected 23 districts in Uganda for P_0=0.1, σ=0.5, and (a) β=0.2, (b) β=0.5, (c) β= 1.7, (d) β = 2.5. The map is colour coded according to the risk of Ebola spreading.*

of infected as well as the severity of EVD. It also reduces the risk of EVD spreading.

EVD spreads from physical contact with bodily fluids such as blood, feces, vomit, saliva, mucus, tears, breast milk, urine, semen, sweat, etc. from infected persons. Therefore, a susceptible person can only be infected with EVD if he or she comes in contact with these bodily fluids from an infected human. So, if infected people are identified and moved to the quarantine, and are restricted from travel, then the infection can be contained. However, non-hospitalized people during early symptomatic stages keep moving for their daily lives and spread the infection to people they come in contact with. Early detection in countries where people are not very health conscious is challenging. Therefore, when an infected case is found, the spread can be contained if the human movement is reduced significantly. However, this is not practical as human movement cannot be controlled.

The activity rate (σ) can be translated to human movement within the network. There-

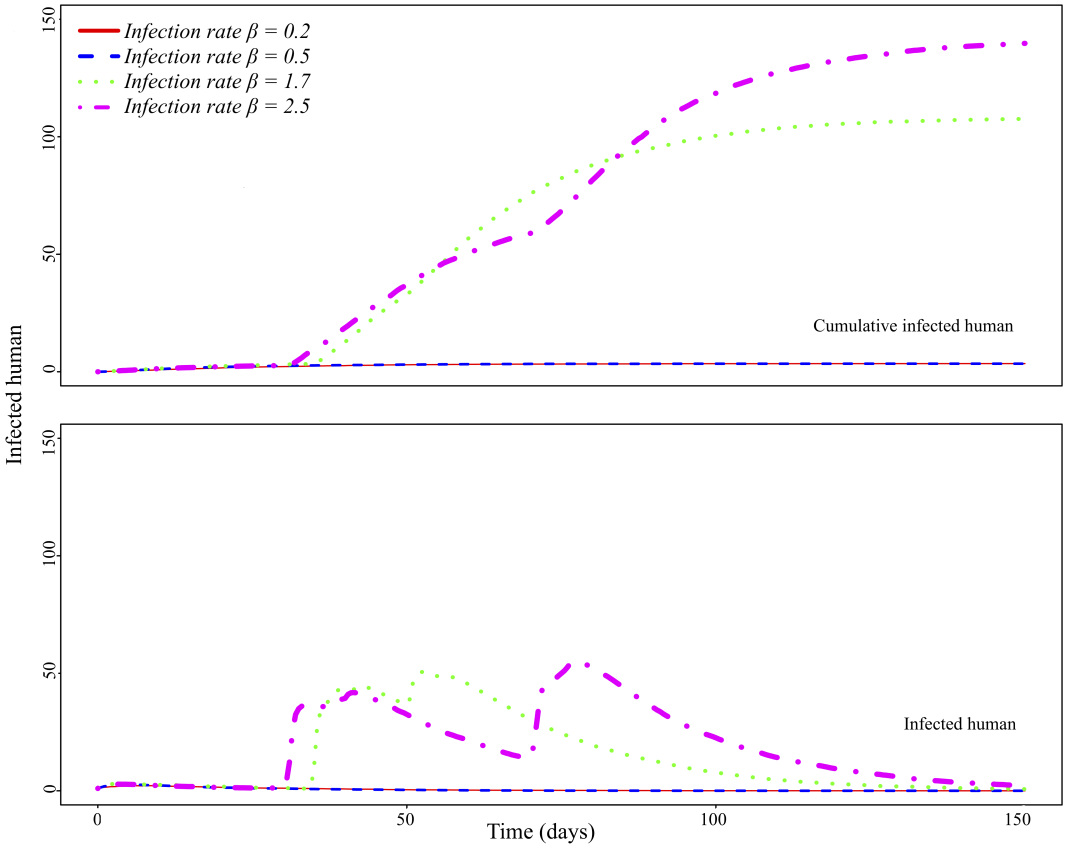


Figure 5.11: *(Top) Average number of cumulative infected and (bottom) average number of infected humans in the Uganda Ebola network for P_0=0.1, σ=0.1 —four different colors represent the average number of infected humans for four values of β. Red, blue, green, and magenta lines correspond to the number of infected humans for β = 0.2, 0.5, 1.7, and 2.5, respectively.*

fore an increasing activity rate means an increased but shorter duration of human movement. From all simulation results, it is evident that an increase in activity rate increases epidemic size as well as the speed of the infection to reach its peak. Therefore, an increasing activity rate means severe and fast-spreading EVD. Simulation results suggest that short duration but frequent human movement results in a greater number of infected humans and a higher risk of EVD spreading than longer duration but a less frequent movement.

The probability at which infection spreads to susceptible people via a potential contact (P_0) has obvious impacts on epidemic size. Increasing P_0 increases epidemic size irrespective of the activity rate σ, which is evident from Figures 5.5, 5.7, 5.9, and 5.11. Comparing Figure 5.5 and Figure 5.9, as well as Figure 5.7 and Figure 5.11, for similar values of the activity rate,

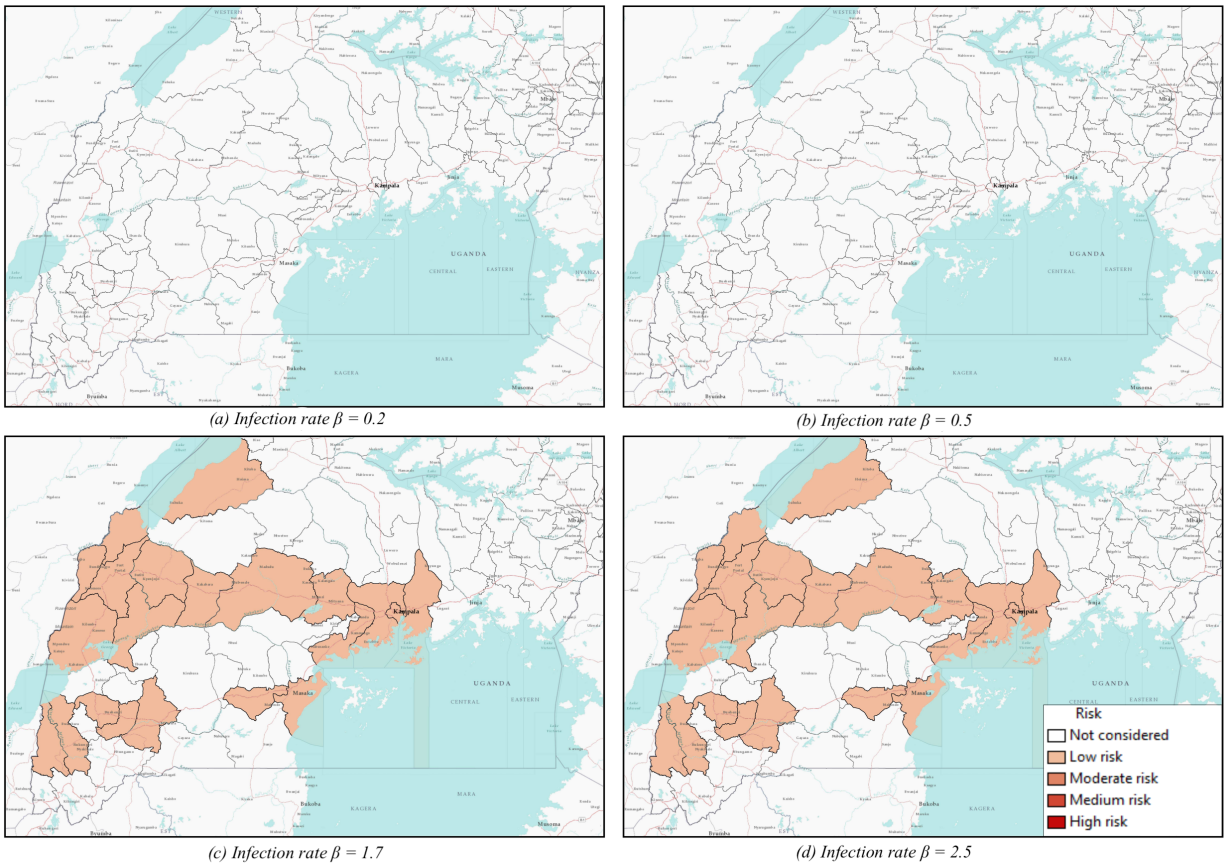


Figure 5.12: *Risk map of Ebola spreading within selected 23 districts in Uganda for P_0=0.7, σ=0.5, and (a) β=0.2, (b) β=0.5, (c) β= 1.7, (d) β = 2.5. The map is colour coded according to the risk of Ebola spreading.*

decreasing P_0 decreases the number of infected humans significantly. P_0 can be translated to the probability of physical contact of infected mobile humans with others. Therefore, our simulation results conform to another mitigation strategy against EVD spreading, which is staying away from contact with people whose status for EVD is unknown. The lower value of P_0 can be achieved by minimizing physical contact among people during movement/travel. If people are aware of the risk of EVD spreading in their area and keep themselves from physical contact with others, it will significantly reduce infected cases if there is an outbreak.

Risk maps show some districts, as shown in Table 5.3 and 5.4, are at higher risk of Ebola spreading in our specific network scenario. The risk assessment provides important information for Ebola preparedness, which includes setting up medical facilities as well as employing different preventive measures against disease spreading. However, resources are

limited, and it is always essential to find a way to utilize resources fruitfully. Our risk only showed examples of our developed risk assessment method, where we have used very limited generalized movement data. Although the proposed risk assessment method can provide guidelines for public health people, it requires more accurate movement data to be used practically in the field. Future work will include an evaluation of the impact of major EVD intervention pillars (i. e., coordination, vaccination, surveillance, risk communication, case management, and safe burials) on risk assessment, to provide a realistic picture under multiple scenarios.

5.4 Conclusions

We present a novel method for risk assessment based on a two-layer temporal network. This method can assess the risk of EVD spreading when accurate network data is available and can be an important tool for public health people during an outbreak. We demonstrate an application of our developed method using a two-layer temporal network formulated with generic and incomplete movement data in Uganda. Simulation results from this two-layer temporal network confirm that reduced physical contact with people while traveling, as well as taking other preventive measures, decreases the risk of EVD spreading. Simulations also show some districts are at higher risk than others in the scenario considered. The identification of the risk provides public health personnel direction for prioritizing their efforts to limit EVD spreading during an outbreak. However, assessed risks are crucially dependent on the network structure and can only be fully trusted for resource allocation once accurate, individual-level movement data in time and space are available.

Chapter 6

Risk assessment of vector-borne disease transmission using spatiotemporal network model and climate data

6.1 Introduction

In modern times, threats posed by infectious diseases have become a significant public health concern due to the increasing connectivity[126]. Emerging infectious diseases (e.g., H7N9, H5N1, and Ebola) as well as endemic diseases (e.g., dengue, chikungunya, measles), pose a severe threat to human health and life[127]. Some infectious diseases have high mortality and morbidity rates (Ebola), and some of these diseases lack treatments or vaccines (dengue)[119;128]. Infectious diseases are creating pandemics due to globalization. For example, 2019 had experienced major dengue outbreaks in many countries in the world, including southeast Asia and Latin America. Other emerging and endemic infectious diseases had also shown to spread rapidly across the globe in recent times. Therefore, accurate risk assessment of disease outbreak is very important for preparedness in this modern world. Risk has been

defined as disease development probability within an individual in a specified time interval by medical epidemiologists and health organizations[129]. Risk is also defined as the potential adverse consequences of unwanted phenomena (disease/event)to human life, health, property, or the environment[130]. Accurate risk assessment models have the potential to improve epidemic prevention and control capabilities.

Due to the complexity in vector-borne diseases spreading, often risk is associated with vector or host suitability, the basic reproduction number, vectorial capacity, vector prevalence, or incidence history[129;131–133]. One method to assess the risk is by detecting disease outbreaks from surveillance data— a retrospective approach, which has a limitation of allowing enough time for preparedness[134]. However, some areas have limited resources for the surveillance system. Especially in developing countries, collected data may not be adequate as most people chose not to use medical facilities unless they have severe conditions. Therefore, disease surveillance data-dependent risk assessment is not always efficient with unreported or under-reported incidences. Therefore, researchers have developed other risk assessment methods to overcome the problem with retrospective methods. The impact of climate change on the vector survival, suitability and pathogen transmission has been assessed for vector-borne diseases in numerous research[135–143]. Unfortunately, very limited researches have included spatial and temporal heterogeneity of the weather conditions, population demography, and movement information in the risk assessment models.

In this context, we develop a risk assessment framework incorporating the aforementioned significant elements for vector-borne diseases. In this chapter, we focus on the mosquito-borne diseases to demonstrate the risk assessment framework as they comprise the majority of vector-borne diseases. This work has three major contributions. The first contribution is the formulation of a spatiotemporal network-based risk assessment framework by incorporating climate data and demographic information, especially for regions with unreported or under-reported incidence data. The second contribution is deriving a spatiotemporal suitability map of competent mosquito species in the disease transmission with only temperature data. The third contribution is the development of spatial risk maps for disease transmission, showing the relative risk of each location compared to others. Additionally, the identification

of the significant-incidence window and peak incidence period is performed by comparing simulation results with data (when available). A serotype analysis is conducted to identify contributing factors for the year-to-year difference in incidence data. This novel risk assessment method is capable of incorporating both human movement and contact patterns as well as impacts of weather factors in human-mosquito interaction.

Finally, an application of the novel framework is presented for dengue spreading in Bangladesh. A spatiotemporal network is developed for human movement in Bangladesh using demographic information, one-month, and two-month lagged climate (temperature and rainfall) data. A map for the spatiotemporal suitability of human-mosquito interaction as well as spatial dengue transmission risk maps are obtained from simulation results. Simulation results matches closely with the significant-incidence window and peak incidence period with Bangladesh dengue transmission dynamics. The year-to-year data variability shows a correlation with the dominant serotype. The combined knowledge obtained from the framework (i.e., significant-incidence window, the peak incidence period, risk map, and spatiotemporal suitability map) provides a guideline to public health personnel in prioritizing spatiotemporal resource allocation to reduce/prevent dengue transmission. Risk maps are developed incorporating generalized human movement data in the spatiotemporal network and have the adaptability to include actual and accurate movement data.

6.2 Materials and method

6.2.1 Risk assessment framework

Our novel risk assessment framework couples a spatiotemporal network-based approach with a compartmental disease model and a spatiotemporal spreading algorithm. The risk assessment framework has five different components. They are as follows-

- Compartmental model
- Pathogen transmission model with climate data

- Spatiotemporal network
- Spatiotemporal spreading algorithm
- Risk calculation

Each of these components is described in subsequent parts of this section.

Compartmental model

Compartmental models express transitions of the host population from one disease state/compartment to another[10]. These compartments are, for example, susceptible, exposed, infectious, recovered, removed, vaccinated, and alert. Some parameters govern inter-compartmental transitions. The transition rate from susceptible to exposed/infected compartment (infection rate) is the most crucial parameter for the vector-borne disease model. The infection rate is correlated with climate/weather dependent factors such as vector abundance and host-vector interactions. Therefore, we incorporate climate data in the infection rate for the developed risk assessment framework.

Pathogen transmission model with climate data

The infection rate for vector-borne diseases has a complicated relationship with the environment and the host. For example, mosquito abundance and their interaction with the host population cause the transition from susceptible to infected (or exposed) states. When the mosquito population is the vector for disease, temperature, and rainfall data are used to develop the correlation between mosquito abundance and their interaction with the host population. This relation can be expressed as *vectorial capacity*— a parameter governing the spread of infection from an infected to a susceptible host via vectors. Vectorial capacity is given as

$$V_c = \frac{ma^2 b_h b_m e^{-\mu_m n}}{\mu_m} \tag{6.1}$$

where the vector parameters used are 1) the average daily vector biting rate (a), 2) the probability of vector to human transmission per bite (b_h), 3) the probability of human to vector infection per bite (b_m), 4) the duration of the extrinsic incubation period (n), 5) the vector mortality rate (μ_m), and 6) mosquito vector density with respect to the host (m)[144–146]. These parameters are specific for the mosquito species and the concerned disease. We choose *Aedes aegypti*, one of the most competent mosquito species in transmitting dengue, Zika, chikungunya, yellow fever, and other severe diseases to model the infection rate/vectorial capacity. Except for mosquito vector density with respect to the host, all other parameters in equation 6.1 can be calculated empirically using spatiotemporal temperature data[146]. The following empirical formulas are used to calculate temperature-dependent parameters, where T is the temperature in degree Celsius.

1) **Biting rate (a):** Liu-Helmersson et al. and Scott et al. developed the following empirical equation from numerous experimental data to model the relationship between temperature and average blood meal frequency of female *A. aegypti*[146;147].

$$a(T) = 0.0043T + 0.0943(day^{-1}) \tag{6.2}$$

2) **Probability of vector to human transmission per bite (b_h):** The empirical equation for the probability of human infection was expressed with the following thermodynamic function[146;148].

$$b_h = 0.001044T(T - 12.286)\sqrt{32.461 - T} \tag{6.3}$$

for $(12.286°C < T < 32.461°C)$

3) **Probability of human to vector infection per bite (b_m):** Lambrechts et al. derived the relationships between temperature and the probability of infection based on empirical data for several *A. aegypti*-borne diseases.[149].

$$b_h = \begin{cases} 0.000729T - 0.9037 & (12.4^\circ C < T \leq 26.1^\circ C) \\ 1 & (26.1^\circ C \leq T < 32.5^\circ C) \end{cases} \tag{6.4}$$

4) **Duration of the extrinsic incubation period (n):** An exponential function was used to fit experimental data for the extrinsic incubation period[149;150].

$$n(T) = 4 + e^{5.15 - 0.123T} \tag{6.5}$$

5) **Vector mortality rate (μ_m):** Yang et al. developed a 4^{th} order polynomial equation fitting experimental data to model the mortality rate with temperature[35].

$$\mu_m(T) = 0.8692 - 0.1590T + 0.01116T^2 - 3.408 * 10^{-4}T^3 + 3.809 * 10^{-6}T^4 \tag{6.6}$$

6) **Mosquito vector density with respect to the host (m):** The mosquito vector density with respect to the host mostly depends on the rainfall. Therefore, in this article, this parameter is expressed as proportional to the weekly average rainfall[151]. We have normalized the average weekly rainfall in each location before using it in the vectorial capacity model.

Spatiotemporal network

To account for the spatiotemporal heterogeneity of the disease transmission risk with changing weather conditions, we propose a spatiotemporal network. The spatiotemporal network is developed using host demographic information such as population density, distribution, and movement. In this network, nodes represent individuals within spatially homogeneous locations, and links represent movements within and between these locations. The network is spatially explicit and has multiple temporal realizations to represent heterogeneities in weather conditions with time and space. This network is then combined with a spatiotemporal spreading algorithm to simulate the spatiotemporal transmission of the infection. The network is periodically updated to reflect the changing weather conditions in the spatiotemporal spreading algorithm.

Spatiotemporal spreading algorithm

Nodes influence each other through statistically independent pairwise interactions in most network-based models. Sahneh et al. developed the *generalized epidemic modeling framework*(GEMF) for stochastic spreading processes over complex networks based on these independent pairwise interactions[29;30]. GEMFsim tool was later developed for numerical simulation of GEMF-based models by implementing the Gillespie algorithm[29]. The combined state of all nodes in a network can be described as a random variable $X_N(t) = [x_1(t), x_2(t), ..., x_i(t)]$, where $x_i(t)$ is the state (compartment) of node i at time t. The transition time from one state to another is expressed as an exponential distribution with a transition rate $\sigma_n(x_n \rightarrow J)$, where J is the destination state after the transition. This transitions can be node-based (dependent only on the node state $x_i(t)$) or edge-based (dependent on the combined network state $X_N(t)$). After a transition occurs, the combined network state will change, and therefore edge-based transition rates will change. However, node-based transition rates remain constant. GEMFsim accounts for changes in transition rates due to the change in the combined network state. However, GEMFsim does not account for the temporal variation of the transition rates due to external factors (weather conditions or human activities).

The temporal variability of the transition rate is very crucial for simulating vector-borne disease transmission. Therefore, it is required to adapt the Gillespie algorithm in GEMFsim to account for the changing rates. In this article, we incorporate the non-homogeneous Gillespie algorithm in the GEMFsim, which works for exponential event distributions and non-constant transition rates[119;152;153]. The modified spreading algorithm is capable of periodically changing transition rates to reflect the temporal heterogeneity of vector-borne disease transmission.

Risk calculation

As the spreading process in our spatiotemporal network is highly stochastic, we need to perform an adequate number of simulations. We keep track of each node's status and count

the numbers of simulations in which a particular node is infected. This count is later used to calculate the risk of the spatial disease spreading. The formula for risk calculation is

$$Risk_j = \frac{\sum_{n=1}^{N_j} I_n}{N_{simulation}} \tag{6.7}$$

where $Risk_j$ is the spreading risk in location j, I_n is the number of simulations where node n is infected, $N_{simulation}$ is the total number of simulations, N_j is the number of nodes (individuals) in j^{th} location. The calculated risk is normalized for comparison with the risk of different spatial locations.

6.2.2 Application of the risk assessment framework for Bangladesh dengue incidence

Dengue is transmitted by *Aedes* mosquitoes in tropical and subtropical regions and may cause a wide range of manifestations from asymptomatic infections to deaths[154]. Arbovirus transmission is known to be driven by the interplay of sex, age, and travel of individuals. Additionally, the transmission also depends on the type of host community (urban/rural), mosquito abundance, and the use of mosquito control measures[155;155–157]. Understanding the relative importance of these factors is required for assessing the risk of dengue accurately. There has been a recent dengue outbreak in Bangladesh with a record number of cases, drawing close attention to assess the transmission risk. Bangladesh has a history of dengue incidence dated back to 1960, and a major outbreak occurred in 2000[158–160]. Since 2000, the Ministry of Health and Family Welfare of the People's Republic of Bangladesh started recording clinical cases, which are reported annually. Recent urbanization throughout the tropical world has accelerated dengue spreading as*Aedes aegypti*— primary vector for dengue transmission— lives in densely populated human-made environments[161]. Bangladesh is a densely populated country with rapid urbanization, which provides a conducive environment for mosquito populations. Therefore, a risk assessment tool for dengue transmission has

become very important in Bangladesh.

Several studies have used the historical time series data for reported dengue incidence. However, getting accurate data on dengue infection is surprisingly difficult. Only 11–32% infected people are likely to have symptoms with just a few being sick enough to require formal medical care[161;162]. Misdiagnosis and under-reporting are common for cases requiring medical care as well. Therefore, risk assessments based on clinical case counts are not always useful, and may just reflect differences in access to healthcare, diagnostics, and the ability to report cases[163]. Record keeping also requires significant resources, which are not always available in Bangladesh. Therefore, our developed framework, which does not require incidence data, can be a useful tool for assessing the spatial dengue transmission risk and the spatiotemporal suitability in Bangladesh. The adapted framework for Bangladesh dengue spreading is presented below.

Compartmental model for dengue

When dengue virus enter into the bloodstream of a susceptible person via infected mosquito bites, the individual becomes exposed to the disease. After a specific time for viral replication, the exposed individual becomes infectious. The infectious individual finally transitions to the removed state after recovery or death. Therefore, there are four specific phases/states concerning the disease. These states are named as susceptible, exposed, infected, and removed, and the model is called *SEIR*. The inter-compartmental transitions are independent Poisson processes with transitions rates expressed in equations 6.8-6.10.

$$Pr[x_i(t+\Delta t)=2|x_i(t)=1,X_N(t)]=\beta_i(t)Y_i\Delta t+o(\Delta t) \tag{6.8}$$

$$Pr[x_i(t+\Delta t)=3|x_i(t)=2,X_N(t)]=\delta\Delta t+o(\Delta t) \tag{6.9}$$

$$Pr[x_i(t+\Delta t)=4|x_i(t)=3,X_N(t)]=\gamma\Delta t+o(\Delta t) \tag{6.10}$$

In these equations, $x_i(t+\Delta t) = 1$, 2, 3, and 4 express the probability of node i occupying the susceptible, exposed, infected or removed state at time $(t + \Delta t)$, respectively. $X_N(t)$ is the combined network state at time t. The transition rate from susceptible to exposed state is an edge-based transition, which is also time-variant due to its dependency on weather conditions. We express this time-variant parameter $\beta_i(t)$ with vectorial capacity for node i at time t, which is calculated from spatiotemporal weather conditions. Y_i is the set of infected neighbors of node i within the spatiotemporal network at time t. The parameter δ is the intrinsic incubation rate, which governs the transition from exposed to infected state. The transition from infected to removed state is expressed with the removal rate γ. Incubation rate δ and removal rate γ is node-based transition rates, whose values are assumed equal to 0.17 and 0.14 respectively, and are time-invariant in this work[115;164]. These values of δ and γ reflect the means of exponentially distributed parameter values used in the spatiotemporal spreading process.

Pathogen transmission model with Bangladesh climate data

Infection rate β is modeled with weekly average temperature and rainfall data in Bangladesh. Climate data are collected from CLIMATE-DATA.ORG for Bangladesh[165]. The upazila level spatial unit is used in this work for network development. Climate data are used to calculate each parameter in the vectorial capacity equation 6.1. All these parameters, except for mosquito vector density with respect to the host, are calculated from the weekly temperature data. Mosquito vector density parameter is assumed proportional to the weekly average rainfall, and a proportional constant is assumed to reflect a realistic outbreak scenario in Bangladesh.

An urbanization factor is assumed to reflect the suitability of *Aedes* mosquito habitat for dengue transmission. The population density is used to classify the urbanization level of each location. Three urbanization factors are used to reflect backcountry (population density $<$ 1000 per square kilometer), rural (1000 $\geq$ population density $<$ 3000 per square kilometer), and urban (population density $\geq$3000 per square kilometer) locations. The final infection

rate used in simulating dengue transmission is equal to vectorial capacity multiplied by the urbanization factor.

Spatiotemporal network

Developing a network is a crucial part of the risk assessment framework. The spatial structure in Bangladesh is as follows-

- Administrative level 1 — 8 divisions
- Administrative level 2 — 64 districts
- Administrative level 3 — 544 upazila

For network development, we have used the administrative level 3 spatial resolution. Therefore, Bangladesh is divided into 544 spatial locations before creating the network. Population data are collected for each spatial location from the City Population website[166]. The population in each location is scaled by 10,000 to reduce the computational burden during simulations.

We assume an Erdos-Renyi network within each upazila, where links are created with a probability of 0.2. Inter-upazila links are created using an exponential dispersion kernel. We use the kernel function e^{-kD} for link generation, where k is a constant, and D is the distance between the source and destination location. We choose the value of k=0.1 for creating the spatiotemporal network. District-level and division-level human movement, along with the exponential dispersion, are incorporated to reflect the human movement patterns. District-level human movement is incorporated by generating links between the capital city Dhaka and all-district cities. Links are created between Dhaka and all-division cities to include division-level human movements in the network. Figure 6.1 demonstrates a simplified outline of the network.

Bangladesh is a small country, having only 147,500 square kilometers of area. However, there is still some spatial and temporal difference between temperature and rainfall throughout the country. This spatiotemporal weather variation is very important for accurately

Figure 6.1: *A simplified diagram for Bangladesh. Each black circle represents the network within an upazila, while lines between circles express the human movement. Circle sizes are scaled according to the human movement (node degree) for that location. Greater circle size indicates a greater amount of human movement flow.*

representing dengue transmission. We included a spatiotemporal heterogeneity in the created network due to weather patterns, i.e., pathogen infection rates on each link. Literature shows the correlation between a one-month and two-month lagged temperature, rainfall, and dengue occurrence in Bangladesh[167]. The heterogeneity in the weather patterns is reflected on a weekly value of infection rate (β) calculated using both one-month and two-month lagged temperature and rainfall data.

We create two instances of the spatiotemporal network for two different outbreak types:

first, major outbreaks spreading throughout the whole country, and second, minor outbreaks spreading only within major divisional cities. We explicitly incorporate the district-level human movement for major outbreak scenarios with exponential dispersion kernel. Division level human movement is incorporated for minor outbreak scenarios along with the exponential dispersion. Upon developing the spatiotemporal network, we apply the stochastic spreading algorithm for risk assessment.

6.3 Results and discussion

Assessed risk in this work is a combination of weather-dependent spatiotemporal suitability and risk maps from the network-based model. Therefore, we present our results in the following two sections such as-

- Spatiotemporal suitability of dengue transmission in Bangladesh
- Risk maps for dengue transmission in Bangladesh

6.3.1 Spatiotemporal suitability of dengue transmission in Bangladesh

Spatiotemporal suitability expresses the spatial and temporal suitability of the vector (mosquito) survival and functioning in pathogen transmission. When comparing dengue epidemic potential over time and space, it is preferable to use the relative vectorial capacity[146]. Relative vectorial capacity is expressed as the vectorial capacity relative to the vector-to-human population ratio and formulated as $RV_c = \frac{a^2 b_h b_m e^{-\mu_m n}}{\mu_m}$. All parameters are temperature-dependent in the relative vectorial capacity definition. A higher relative vectorial capacity indicates a higher potential for the dengue epidemic. We calculate the relative vectorial capacity for *Aedes* mosquito in Bangladesh to infer the suitability of dengue spreading. There exists a threshold value for relative vectorial capacity beyond which *Aedes* mosquitoes can function properly in transmitting dengue infection. A relative vectorial capacity greater than 0.6 indicates the suitability of dengue spreading[146]. The spatiotemporal suitability maps for dengue spreading in Bangladesh are presented in Figure 6.2 for each month of the year. The

relative vectorial capacity is found higher than the threshold value (0.61) for many months of the year in almost all locations (red and yellow regions).

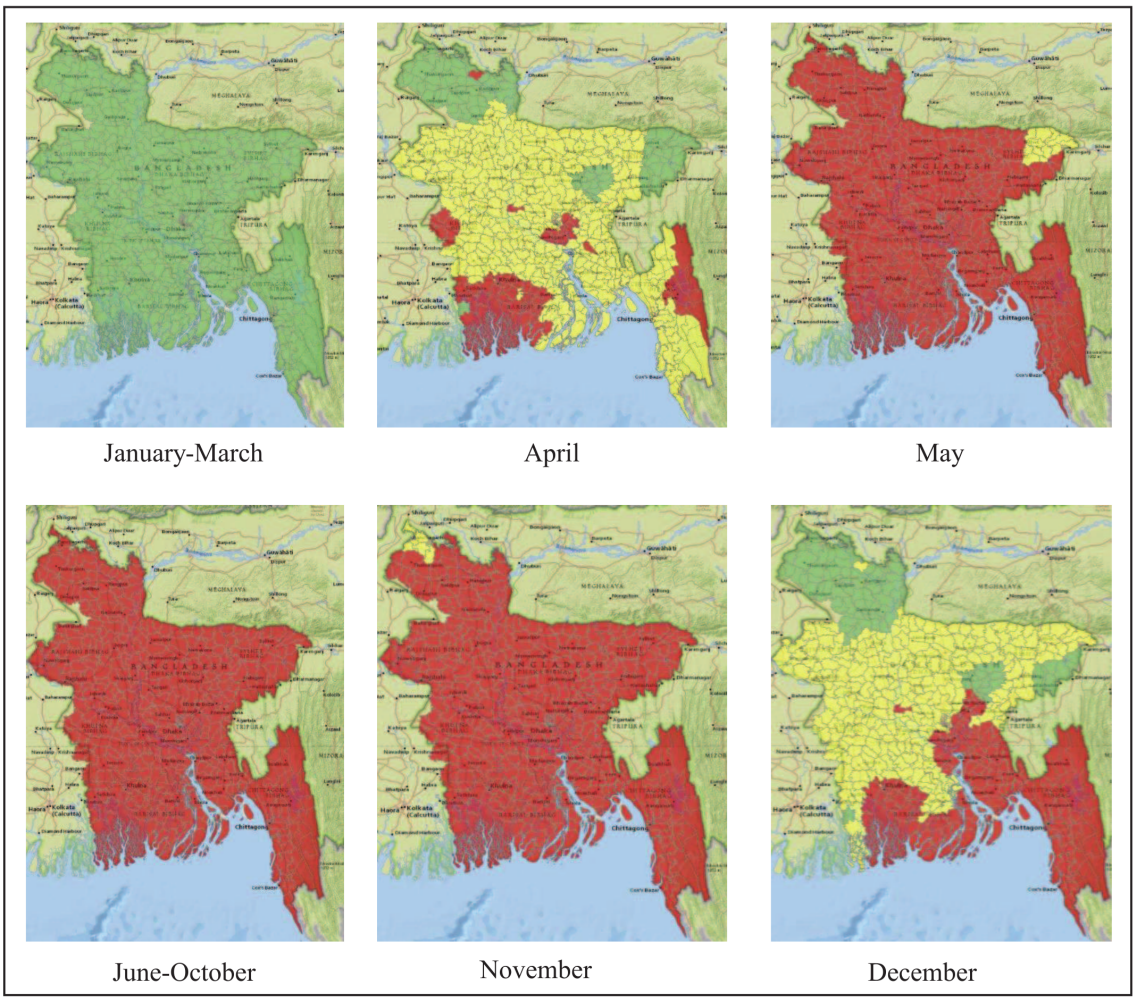


Figure 6.2: *Spatiotemporal suitability maps for dengue transmission in Bangladesh based on temperature. The map represents suitability in the following manner- green (lowly suitable), yellow (moderately suitable), and red (highly suitable.)*

.

It is evident from Figure 6.2 that months between January and March are poorly suited for dengue transmission. In April, some southern parts of the country, as well as capital city Dhaka, become highly suitable. The other parts of the country, except the north most corner, become moderately suitable. Starting from May, the whole country becomes highly suitable, and the situation remains similar until November. The most northern part of the country becomes lowly suitable while the capital and southern part stays highly suitable in

December. After December, the whole country becomes poorly suited again, which continues until April.

Dengue incidence data since 2000 shows cases throughout the year, which supports our results for the suitability of dengue transmission. Bangladesh is a tropical country, which provides a suitable temperature for mosquito survival year-round. However, for temperate regions with widely varying temperatures, mosquitoes may not be able to survive in the coldest months.

6.3.2 Risk maps for dengue transmission in Bangladesh

The network-based model enables us to simulate the spreading process within the spatiotemporal network described above. Every year, the dengue outbreak in Bangladesh shows a different trend. Some years, cases are reported from most parts of the country, while some years outbreaks are mostly limited within divisional cities. Depending on the spreading, outbreaks are divided into two categories— major outbreaks with widespread dengue cases and minor outbreak with cases mostly in some divisional cities. Therefore, two distinct simulation scenarios are assumed to match the two different outbreak types. Simulations are performed with parameters calculated using both one-month and two-month lagged temperature and rainfall data, which are presented in the subsequent parts of this section.

Scenario 1: major outbreak

A major outbreak is defined when dengue cases are widespread throughout the whole country. For major outbreaks, the network is generated with district-level human movements incorporated. Dhaka is the capital of the country; therefore, frequent movements between all district towns are assumed to and from Dhaka. We performed one-thousand iterations for each scenario to account for the stochasticity in the simulation results. Simulations are started with an entirely susceptible population with one infected human in the initial outbreak location. The initial outbreak location is also changed to see the impact of human movements on the transmission risk. Simulation results for one-month lagged climate data,

and two initial conditions— Dhaka and Chittagong outbreak starting are presented in Figure 6.3.

Figure 6.3(a) and 6.3(d) shows the fractions of infected people and cumulative infected people with 95 % confidence interval when the simulation started in Dhaka and Chittagong, respectively. The curve for the average number of infected shows two peaks in Figure 6.3(a) and 6.3(d). The first peak refers to the rapid spreading within the vicinity of the initial outbreak location, and the second peak represents the widespread outbreak. For novel vector-borne diseases, the infections start with a single infected human or a single infected mosquito. If we assume a single infected human started the infection, then the infection will start locally with competent mosquitoes biting the infected person. As mosquitoes have a short flying range, the infection will be local at the beginning, and this is evident from the smaller initial peak. The size of the peak depends on the population size, level of urbanization, and the area of the outbreak location. Being Dhaka densely populated, an outbreak starting in Dhaka will result in a pronounced initial peak due to only the local transmission (Figure 6.3(a)). We started our simulation in April as it is the first month of high suitability for *Aedes* mosquito within a year for Bangladesh. However, the suitability increases with time, as shown in Figure 6.2 and the infection starts spreading to distant locations due to human movement as well as higher suitability. Hence, we have another peak in August with numerous cases countrywide. Although Figure 6.2 shows all locations are highly suitable from May to November, the underlying mosquito dependent infection rate keeps increasing until August and attains higher values in June-August. Therefore, despite human movements from Dhaka to other locations all year, due to the lower value of infection rates, no widespread infections start until June. With the higher value of the infection rate, the number of infected individuals keeps increasing from June and attains its peak in August. Around 16% of the total population in the network becomes infected in this outbreak scenario. Almost 70% of the total infection are observed within the period of July-September. The peak consists of 2.5% of the entire population in our developed network. These infected people during August may require hospital care, which will be a huge burden on the healthcare system. Therefore, proper measures should be taken, and

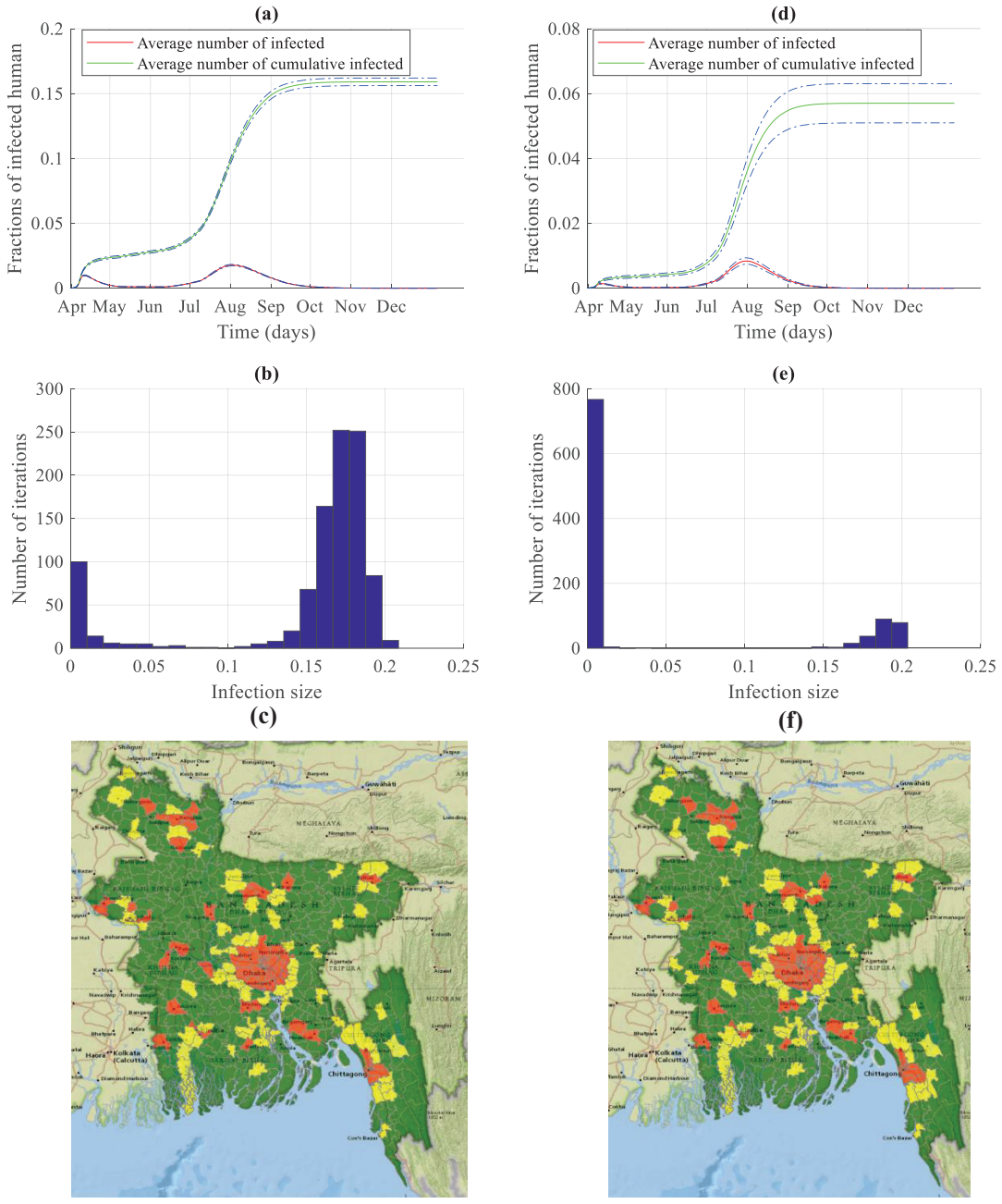


Figure 6.3: *Simulation results and risk maps for dengue transmission in Bangladesh for a major outbreak. The left side panels are results of simulations started in Dhaka, while the right side panels are results of simulations started in Chittagong. Panels (a) and (d) show the dengue transmission dynamics; Panels (b) and (e) present histograms of the number of simulations and infection size; Finally, panels (c) and (f) display risk maps for dengue infection.*

more resources should be allocated for dengue healthcare during July-September. Figure 6.3(d) shows the dynamics when dengue infection started in Chittagong. Total fractions of cumulative infected people and infected people during peak time are both significantly lower for this outbreak scenario, which can be attributed to the fact that Chittagong is not very densely populated as well as not well connected to the whole country as Dhaka. Therefore, both the initial smaller peak and the second higher peak are smaller compared to the peaks in the Dhaka outbreak scenario. When the infection starts in Chittagong, around 6% of the population becomes infected compared to 16% in the Dhaka outbreak scenario. Therefore, the infection starting location is a crucial determining factor of the extent and the epidemic dynamics.

Figure 6.3(b) and 6.3(e) show the histogram with fractions of cumulative infected humans in the Bangladesh dengue network and the number of simulations performed. The x-axis represents fractions of infected humans in simulation, and the y-axis shows the number of simulations where a particular infection size is obtained. Figure 6.3(b) shows that almost 80% simulation results in 10-20% infected humans in the representative network when the initial outbreak happens in Dhaka. This accounts for the narrower confidence interval in Figure 6.3(a). Figure 6.3(e) express around 20% probability of 10-20% human being infected. This variability in the fractions of infected individuals accounts for the wider confidence interval in the simulation results in Figure 6.3(d).

Transmission dynamics presented in Figure 6.3 are obtained using the network with district-level human movement incorporated and one-month lagged temperature and rainfall data. However, when we perform simulations with two-month lagged temperature and rainfall data, the simulation is started in May, and the peaks are a month delayed. The analysis of Bangladesh dengue incidence data since 2000 showed the occurrences of peaks in July-October[158;159]. Therefore, two-month lagged data also show significant similarity with the actual incidence data. Risk maps are similar for both one-month lagged, and two-month lagged data.

Scenario 2: minor outbreak

Dengue infections are often confined within divisional cities, and a widespread outbreak does not happen. Therefore, we propose another scenario where we explicitly incorporate human movement from Dhaka city to other divisional cities along with the distance-based exponential movement kernel. We call this scenario a minor outbreak as the infection does not spread countrywide. We perform simulations for both one-month and two-month lagged climate data.

The results are presented in Figure 6.4, when simulations are performed with one-month lagged climate data and the same initial conditions as major outbreak scenarios, namely starting simulations in Dhaka and Chittagong.

Figure 6.4(a) and 6.4(d) show a similar trend as shown in Figure 6.3(a) and 6.3(d). For one-month lagged data, a major peak is observed in August with a rapidly increasing infection within July- September. However, comparing Figure 6.4 and Figure 6.3, we can see that both the value of the fraction of infected people during the peak infection time and the whole outbreak period are smaller during minor outbreaks.

Human movement can also be considered as a crucial factor in the vector-borne disease spreading, although the pathogen transmission does not happen via direct physical contact. An exposed/infected person may move/travel to a different location and become infectious after reaching the destination. That person can be bitten by a competent local mosquito and may start a local outbreak in the destination location. Therefore, mosquitoes are responsible for the local transmission, while human movement is mostly responsible for long-distance pathogen transmission during a period of higher suitability. The reduction in the number of infections in the minor outbreak scenario can be attributed to the reduced human movement volume within the network than the major outbreak scenario.

It is evident from the comparison between panels (a) and (d) of Figure 6.4 and 6.3 that the major dengue spreading period in Bangladesh is June-September. Finding this time window for significant transmission is very crucial for public health officials. The identification of the significant-incidence window will enable them to take prompt actions during the surge

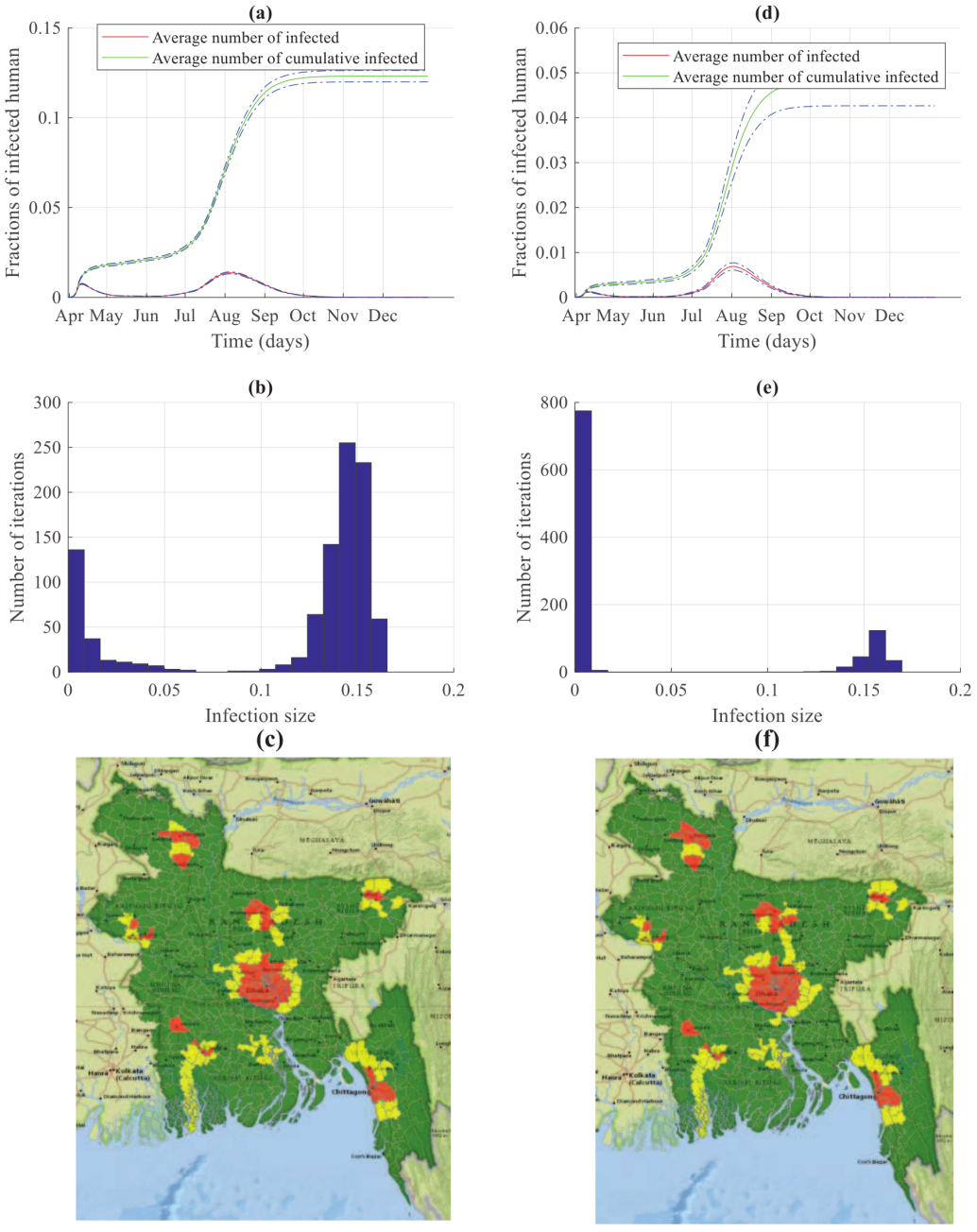


Figure 6.4: *Simulation results and risk maps for dengue transmission in Bangladesh for a minor outbreak. The left side panels are results of simulations started in Dhaka, while the right side panels are results of simulations started in Chittagong. Panels (a) and (d) show the dengue transmission dynamics; Panels (b) and (e) present histograms of the number of simulations and infection size; Finally, panels (c) and (f) display risk maps for dengue infection.*

of infection. The major action for controlling a vector-borne disease is always controlling the vectors. Therefore, control measures need to focus mostly on the significant-incidence months to reduce the mosquito population when resources are inadequate for the whole year.

Figure 6.4(c) and 6.4(f) show the normalized risk maps for dengue spreading during a minor outbreak. It is evident from the figure that high-risk areas are confined within major division cities and their nearby locations in contrast to Figure 6.3(c) and 6.3(f), where many locations throughout the country are at high risk. This reduction in the spreading risk can be attributed to the reduced human movement in the minor outbreak scenario. Therefore, making people aware of the human movement's impact on long-distance travel through social media or radio/TV broadcasting would help contain the epidemic.

Simulations are also performed for two-month lagged climate data, and all the results are similar to the ones obtained with one-month lagged data.

Our proposed framework is a generalized risk assessment tool based on climate and demographic data, which can be used for risk assessment, especially in regions with unreported or under-reported incidence data. Risk maps developed in this work are generated with the generalized concept of human movement within Bangladesh. The framework can incorporate more detailed and accurate human movement data. The incorporation of detailed movement data will provide a more accurate assessment of the transmission risk of each location. Once proper and accurate movement data is incorporated in the network, the control measures should be applied to the high-risk areas first, followed by the medium and low-risk areas depending on the availability of resources.

6.3.3 Serotype analysis

Since 2000, there are dengue cases every year in Bangladesh. However, the number of cases varies from year to year. Importantly, the available data concern only hospitalized and reported cases of dengue. In this section, we show the existence of a correlation between the number of cases and the circulating DENV serotypes. Dengue fever can be caused by any of four genetically related dengue virus (DENV) serotypes (DENV1, DENV2, DENV3,

and DENV4)[168]. After recovering from infection with one dengue serotype, a person has immunity against that particular serotype[168]. Unfortunately, the person can be infected again with any of the remaining three dengue serotypes[169]. Subsequent infections often put individuals at a greater risk for severe dengue illnesses than those who have not been previously infected[170]. A bar graph of yearly dengue incidences in Bangladesh with serotypes is presented in Figure 6.5.

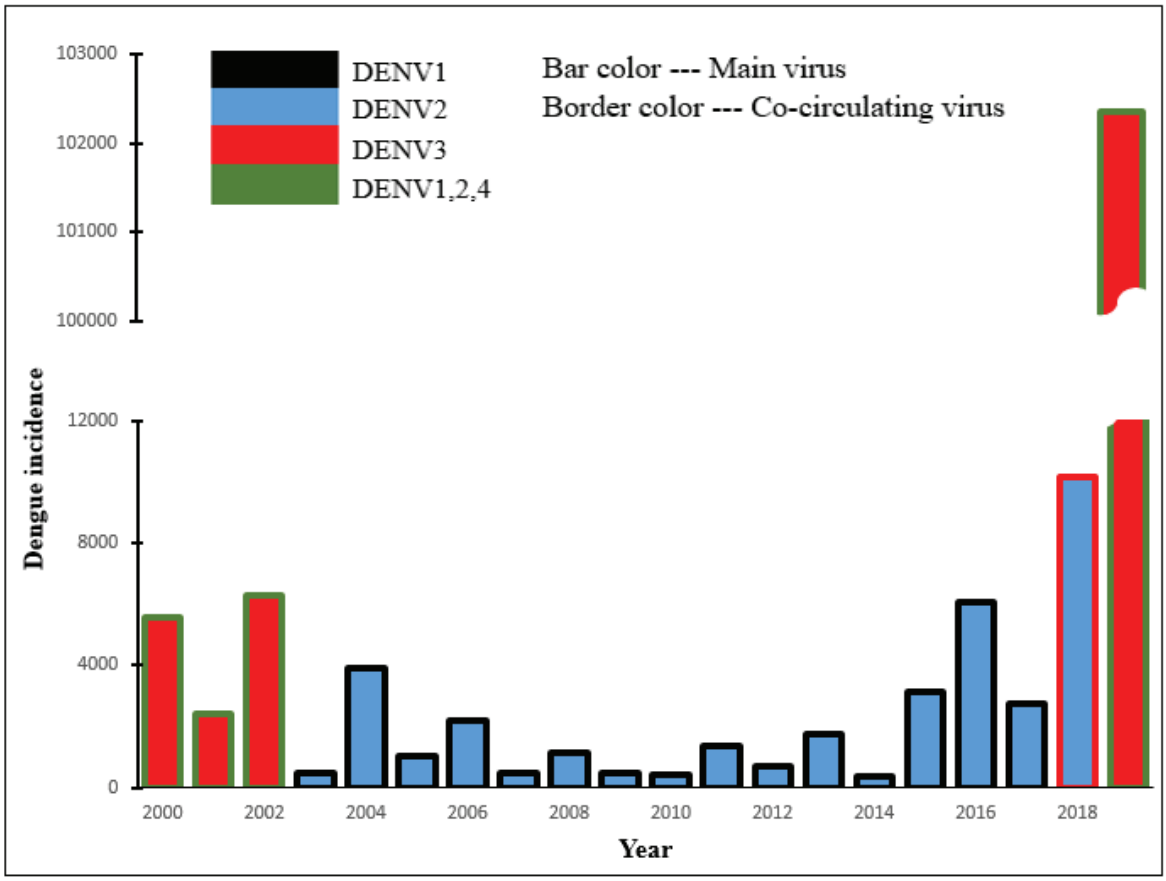


Figure 6.5: *Serotype analysis of dengue spreading in Bangladesh since 2000 The bar chart presents the number of dengue cases with the circulating serotypes each year in Bangladesh. The main bar color represents the dominant serotype, while the border represents other circulating serotypes.*

The primary bar color represents the dominant circulating serotype, while the border-color represents the co-circulating serotypes. The x-axis shows the year, and the y-axis shows the number of reported dengue cases in Bangladesh. The bar and border colors are as follows- black for DENV1, blue for DENV2, red for DENV3, and green for DENV1,2,4

together. The height of each bar corresponds to the number of cases in the corresponding year. Figure 6.5 shows that the dominant circulating serotype was DENV3, with the other three co-circulating, during the years 2000-2002. Within the period 2003-2016, DENV-2 was the dominant serotype, and DENV-1 was co-circulating[169]. The DENV3 reemerging in 2017 resulted in a significant increase in the number of cases during 2018. However, DENV2 was still the dominant circulating serotype in 2018[169]. If a serotype is circulating long enough within a population, all recovered people may become immune to that specific serotype. An increase in the number of cases in 2018 can be attributed to the reemergence of DENV3. DENV3 became the dominant circulating serotype in 2019, which created an extreme and unprecedented surge in the number of infections. More than a hundred thousand cases were reported in 2019— which is more than double of combined cases in the previous nineteen years— due to a vast susceptible population for DENV3 serotype. Currently, DENV3 is already in circulation in Bangladesh. Neighboring countries of Bangladesh have all four serotypes. Therefore, Bangladesh is always at risk for all four serotypes, including DENV4. Healthcare personnel should be vigilant to identify dengue patients before the significant-incidence period to identify the circulating serotypes. The introduction of a new serotype will produce a surge in dengue infections with a high probability.

6.3.4 Peak timing validation

The yearly reported cases varied widely for dengue incidence in Bangladesh. For example, there were 10148 cases in 2018, and more than a hundred thousand cases were reported in 2019[169;171]. Therefore, there is high variability in year-to-year dengue cases. From the serotype analysis in the previous section, this 2019 unprecedented increase can be attributed to the DENV-3 circulation. The yearly number of dengue cases is a complex combination of circulating serotypes, the movement patterns of the human population, and the measures taken for mosquito control. For widely varying year-to-year case numbers, we used the peak incidence time and the transmission dynamics to compare our simulation results with the incidence data. Peak incidence happened mostly in August and September in Bangladesh[169].

Simulation results with one-month lagged climate data show incidence dynamics with the major peak in August. Two-month lagged climate data resulted in a peak in September. We compared our simulation with actual incidence data for 2018 and 2019 in Figure 6.6.

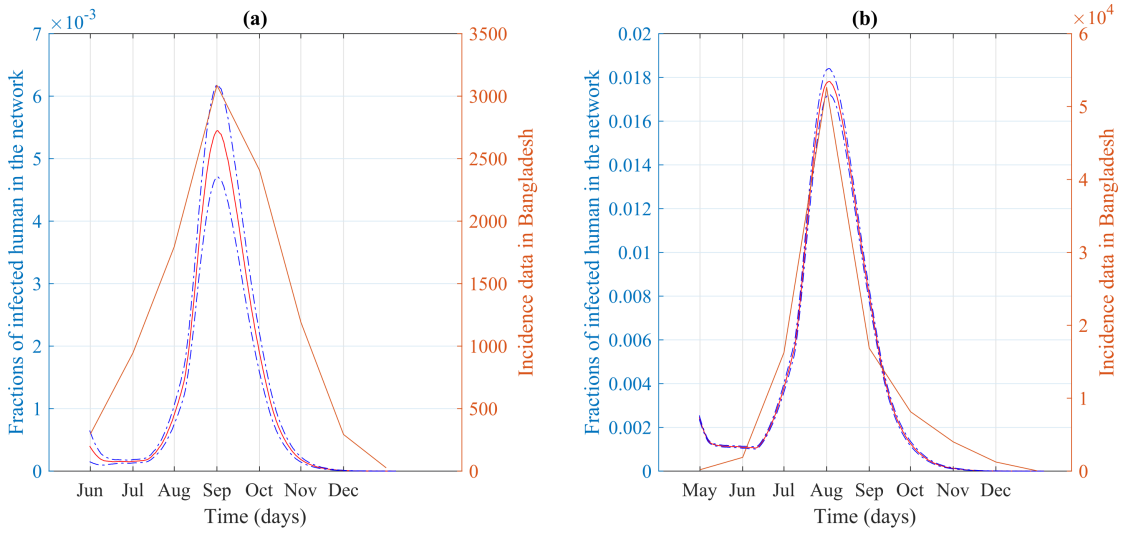


Figure 6.6: *(a) Comparison of peak time from our simulation with incidence data during a minor outbreak in 2018; (b) Comparison of peak time from our simulation with incidence data during a major outbreak in 2019.*

In 2018, the peak of the incidence data was observed in September, as shown in Figure 6.6(a). Simulation results for two-month lagged climate data show a peak in September in accordance with 2018 incidence data in Figure 6.6(a). The disease dynamics from simulations with one-month lagged climate data shows a similar trend with a coinciding peak for 2019 (Figure 6.6(b)).

6.3.5 Application of control measures

The application of control measures decreases the spatial spread as well as the number of dengue cases during an outbreak. Main control measures include spraying insecticide and adopting preventive measures to avoid mosquito bites. To stay away from mosquito bites, one can wear clothes that cover most parts of the body, sleep under bed nets, and use mosquito repellents. Decreasing the probability of mosquito bites per day as well as decreasing mosquito abundance will check the widespread outbreak of dengue. To demonstrate the

impact of the control measures, we performed our simulations with multiple reduced infection rates. Figure 6.7 shows the effect of a 50% reduction in the infection rate in the disease dynamics and the corresponding risk map.

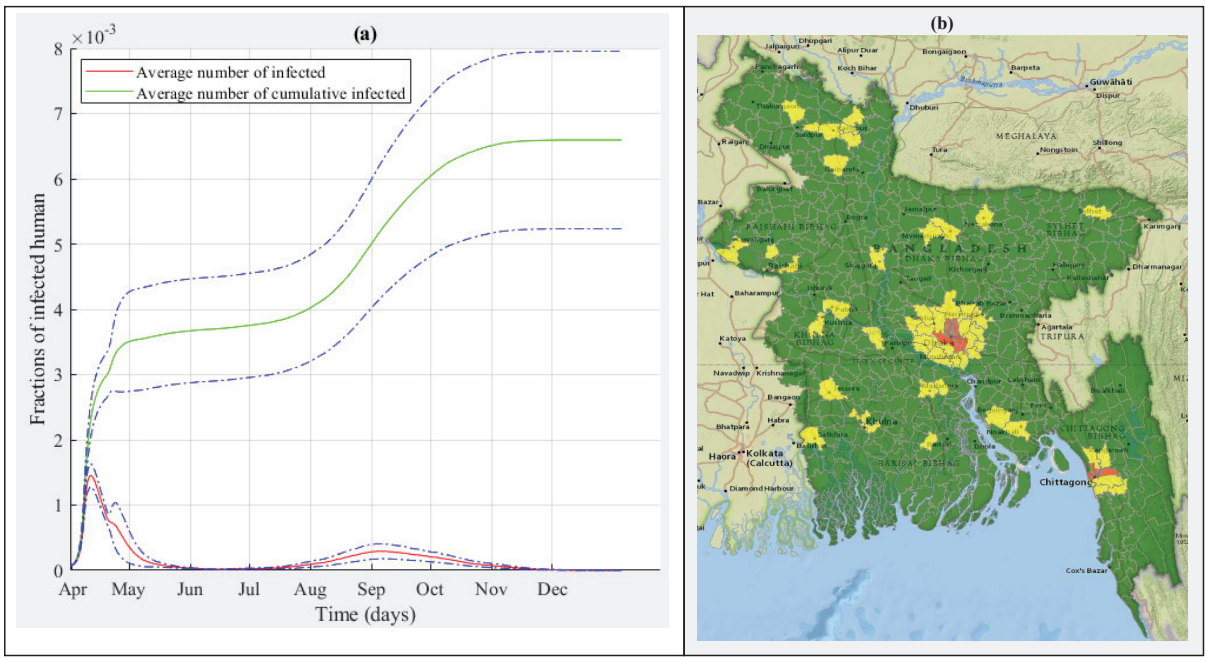


Figure 6.7: *(a) Temporal spreading of dengue with control measures implemented; (b) Spatial risk map of dengue spreading when control measures are applied.*

Figure 6.7(a) shows the disease dynamics after the control measures implemented when the infection started in Chittagong. The fraction of infected individuals is greatly reduced as compared to Figure 6.4 and 6.3. Figure 6.7(b) shows the risk map for dengue transmission after control measures have been applied. The high-risk areas now become confined within the initial outbreak location and some areas in the capital city. The application of control measures reduced both the risk of spatial dengue transmission as well as the total number of cases. Therefore, proper application of control measures, especially during the significant-incidence period identified from our simulation, would be very effective in reducing the epidemic transmission.

6.4 Conclusions

Climate data can be used to develop a generalized risk assessment framework for vector-borne diseases together with demographic data— spatial distribution of individuals and their movement patterns. Therefore, we develop a novel risk assessment framework with a spatiotemporal network model and a non-homogeneous Gillespie algorithm using both climate and demographic data. The assessed risk from this framework is comprised of spatiotemporal suitability maps and spatial risk maps. Spatiotemporal suitability maps show the spatial and temporal suitability of vector-borne transmission. Spatial risk maps represent the disease transmission risk of each location compared to other locations on the map. This framework also identifies the high risk or elevated risk-months as well as the peak incidence period within a year.

Upon development, the framework is applied to the study of dengue transmission in Bangladesh for major and minor outbreak scenarios. The difference between major and minor outbreaks is defined by different levels of human movement to demonstrate the critical role of human dispersal on widespread pathogen transmission. Reduced human migration throughout the country will reduce the infection spread to divisional cities. We generate the spatiotemporal suitability map and the risk maps for Bangladesh dengue transmission. Simulation results also showed a similar significant-incidence window as well as the peak incidence period with reported dengue incidence data in Bangladesh. Serotype analysis indicates the importance of identifying circulating DENV serotypes before the significant-incidence window. The possibility of a major outbreak is associated with the introduction and reemergence of new DENV serotypes. Simulations of control measure applications, such as mosquito control or other preventive behaviors, have shown a significant decrease in the number of cases as well as risk. The proposed risk assessment tool provides guidelines for public health officials to prioritize resource allocation and control measure application according to the estimated risk.

Chapter 7

Summary and future works

7.1 Summary

In this dissertation, we investigate infectious disease models through a network approach to develop a general guidelines for network-based model formulation. We formulate realistic network models for specific disease transmissions and simulate with a stochastic spreading algorithm to help public health policymakers effectively allocate resources and suggesting mitigation measures and risk assessment.

The stochastic spreading algorithm for the network-based models requires disease transmission parameters, which are specific to the disease, spatial location, and the historical incidence data. We develop a parameter estimation framework using a sequential Monte Carlo filter to estimate the parameters. This method is a real-time parameter estimation framework, capable of incorporating the new incidence data upon its availability. We adapt an ensemble Kalman filter (EnKF) for simultaneously estimating the disease transmission parameters as well as for forecasting the number of new cases. Being the ensemble Kalman filter being an online inferential method, it can perform real-time forecasts during an outbreak. The framework is capable of accurate short to mid-term forecasts.

We incorporate the changing contact structures among individuals due to movement with a two-layer temporal network, namely a static and a temporal/dynamic layers. The static

layer represents permanent contacts among individuals, while the temporal layer presents the changing contact structures due to movement. The two-layer temporal network has been combined with a compartmental model and the Gillespie algorithm for assessing the spatial transmission risk of infectious diseases. The final results of the frameworks are spatial risk maps and disease transmission dynamics. The spatial risk maps provide some high-risk locations for public health personnel to focus concerning disease preparedness and mitigation interventions. The disease transmission dynamics with the increasing human movement demonstrates a growing burden in the healthcare facility during the peak incidence period. Therefore, reducing the size of the peak infected human, alternatively known as flattening the peak, is required to provide healthcare facilities to all infected humans.

We model the pathogen transmission from an infectious individual to a susceptible one for vector-borne diseases with the vectorial capacity, which in turn is dependent on the temperature and rainfall data. We developed a risk assessment framework using climate (average temperature and rainfall) and host demographic (host density and movement) data for vector-borne diseases. This framework is particularly suitable for the introduction of vector-borne diseases in a new geographic location or regions with unreported or under-reported incidence data. We also developed a spatiotemporal network that incorporates the spatial and temporal heterogeneity of vectorial capacity. The outcomes of the risk assessment framework are the spatiotemporal suitability map and the spatial risk map. The framework also identifies the significant-incidence window and peak incidence period— two critical factors for public health officers in identifying the time-frame to concentrate the majority of the resource allocation. Both the aforementioned risk assessment framework and the resulting risk maps are crucially dependent on the host movement and can be a useful tool for preparedness with accurate host movement data.

7.2 Future works

We have performed a sensitivity analysis of parameters in Chapter 2 and Chapter 5. The sensitivity analysis's primary goal is to demonstrate the impact of crucial change in the host,

vector, and environmental factors in the disease transmission dynamics. The sensitivity analysis satisfies the goal mentioned above of our network-based models, while calls for a further specification in the parameters. In Chapter 3, we have shown the parameter estimation framework using historical incidence data. However, the parameter estimation framework needs to be incorporated within the network-based for the spatial location to provide more accurate mitigation measures. Once the parameters are estimated for a specific model, then the network-based model's simulation results will be more useful for public health personnel. However, the main challenge in the inclusion of the parameter estimation method in the network-based simulation framework was the unavailability of accurate historical incidence data. Another challenge is the high sensitivity of the filtering methods to the initial conditions, which can be addressed with recently developed modified filtering methods. This direction of future research will be fruitful, advancing the application of network-based or in general epidemic models for practical purposes.

In Chapters 5 and 6, we have proposed risk assessment frameworks for disease transmission using network models. Although we applied the method for a specific location, we expect this framework can be successfully applied to other locations. The network structure— an element of crucial impact on the assessed risk, is variable depending upon the location. Therefore, extending these risk assessment methods to other spatial locations with detailed host demographic data can provide guidelines in risk assessment, and control of current and potential outbreaks.

Bibliography

[1] Henri Berestycki, Samuel Nordmann, and Luca Rossi. Modeling propagation of epidemics, social unrest and other collective behaviors. *arXiv preprint arXiv:2005.09865*, 2020.

[2] Julie Simmons Ivy, Muge Capan, Karen Hicklin, Nisha Nataraj, Irem Sengul Orgut, Amy Craig Reamer, and Anita Vila-Parrish. To be healthy, wealthy, and wise: Using decision modeling to personalize policy in health, hunger relief, and education. In *Women in Industrial and Systems Engineering*, pages 233–274. Springer, 2020.

[3] Constantinos I Siettos and Lucia Russo. Mathematical modeling of infectious disease dynamics. *Virulence*, 4(4):295–306, 2013.

[4] PK Tapaswi, Asit K Ghosh, and BB Mukhopadhyay. Transmission of japanese encephalitis in a 3-population model. *Ecological modelling*, 83(3):295–309, 1995.

[5] Ronald Ross. An application of the theory of probabilities to the study of a priori pathometry. part i. *Proceedings of the Royal Society of London. Series A, Containing papers of a mathematical and physical character*, 92(638):204–230, 1916.

[6] Ronald Ross and Hilda P Hudson. An application of the theory of probabilities to the study of a priori pathometry. part ii. In *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, volume 93, pages 212–225. The Royal Society, 1917.

[7] Caterina M Scoglio, Claudio Bosca, Mahbubul H Riad, Faryad D Sahneh, Seth C Britch, Lee W Cohnstaedt, and Kenneth J Linthicum. Biologically informed individual-based network model for rift valley fever in the us and evaluation of mitigation strategies. *PloS one*, 11(9):e0162759, 2016.

[8] Annelise Tran, Carlene Trevennec, Julius Lutwama, Joseph Sserugga, Marie Gely, Claudia Pittiglio, Julio Pinto, and Veronique Chevalier. Development and assessment of a geographic knowledge-based model for mapping suitable areas for rift valley fever transmission in eastern africa. *PLoS neglected tropical diseases*, 10(9), 2016.

[9] Sifat Afroj Afroj Moon, Tanvir Ferdousi, Adrian Self, and Caterina Scoglio. Estimation of swine movement network at farm level in the us from the census of agriculture data. *bioRxiv*, page 488767, 2018.

[10] Matt J Keeling and Ken TD Eames. Networks and epidemic models. *Journal of the Royal Society Interface*, 2(4):295–307, 2005.

[11] Qiong Liu and Jianhua Yang. Application of hidden markov model on the prediction of hepatitis b incidences. In *International Conference on Smart Health*, pages 209–221. Springer, 2019.

[12] C Souty, R Jreich, Y Le Strat, C Pelat, PY Boëlle, C Guerrisi, S Masse, T Blanchon, T Hanslik, and C Turbelin. Performances of statistical methods for the detection of seasonal influenza epidemics using a consensus-based gold standard. *Epidemiology & Infection*, 146(2):168–176, 2018.

[13] Yilan Liao, Bing Xu, Jinfeng Wang, and Xiaochi Liu. A new method for assessing the risk of infectious disease outbreak. *Scientific reports*, 7:40084, 2017.

[14] Adewara Johnson Ademola and Gbadegeshin Monsurat. Modeling tuberculosis (tb) using higher order markov model. *Covenant Journal of Physical and Life Sciences*, 6 (2), 2019.

[15] Tru Hoang Cao, Anh Duy Nguyen, Tuan Quang Dinh, Quang Chan Luong, and Hai Thanh Diep. Forecasting dengue incidences: Statistical and dynamic models. *Oxford Journal of Intelligent Decision and Data Science*, 2018.

[16] Amanda A Koepke, Ira M Longini Jr, M Elizabeth Halloran, Jon Wakefield, and

Vladimir N Minin. Predictive modeling of cholera outbreaks in bangladesh. *The annals of applied statistics*, 10(2):575, 2016.

[17] Dianbo Liu, Leonardo Clemente, Canelle Poirier, Xiyu Ding, Matteo Chinazzi, Jessica T Davis, Alessandro Vespignani, and Mauricio Santillana. A machine learning methodology for real-time forecasting of the 2019-2020 covid-19 outbreak using internet searches, news alerts, and estimates from mechanistic models. *arXiv preprint arXiv:2004.04019*, 2020.

[18] Salah Ghamizi, Renaud Rwemalika, Maxime Cordy, Yves Le Traon, and Mike Papadakis. Pandemic simulation and forecasting of exit strategies: Convergence of machine learning and epidemiologicalmodels. Technical report, University of Luxembourg, 2020.

[19] Biobele J Brown, Alexander A Przybylski, Petru Manescu, Fabio Caccioli, Gbeminiyi Oyinloye, Muna Elmi, Michael J Shaw, Vijay Pawar, Remy Claveau, John Shawe-Taylor, et al. Data-driven malaria prevalence prediction in large densely-populated urban holoendemic sub-saharan west africa: Harnessing machine learning approaches and 22-years of prospectively collected data. *arXiv preprint arXiv:1906.07502*, 2019.

[20] Tanujit Chakraborty and Indrajit Ghosh. Real-time forecasts and risk assessment of novel coronavirus (covid-19) cases: A data-driven analysis. *Chaos, Solitons & Fractals*, page 109850, 2020.

[21] Stephen Parnell, Frank van den Bosch, Tim Gottwald, and Christopher A Gilligan. Surveillance to inform control of emerging plant diseases: an epidemiological perspective. *Annual review of phytopathology*, 55:591–610, 2017.

[22] Fred Brauer. Compartmental models in epidemiology. In *Mathematical epidemiology*, pages 19–79. Springer, 2008.

[23] Romualdo Pastor-Satorras, Claudio Castellano, Piet Van Mieghem, and Alessandro

Vespignani. Epidemic processes in complex networks. *Reviews of modern physics*, 87 (3):925, 2015.

[24] Sifat Afroj Moon, Lee W Cohnstaedt, D Scott McVey, and Caterina M Scoglio. A spatio-temporal individual-based network framework for west nile virus in the usa: spreading pattern of west nile virus. *bioRxiv*, page 438366, 2018.

[25] Aram Vajdi, David Juher, Joan Saldana, and Caterina Scoglio. Sis epidemics in multilayer-based temporal networks. *arXiv preprint arXiv:1812.05073*, 2018.

[26] Matt J Keeling and Pejman Rohani. *Modeling infectious diseases in humans and animals*. Princeton University Press, 2008.

[27] Matt J Keeling and Pejman Rohani. *Modeling infectious diseases in humans and animals*. Princeton University Press, 2011.

[28] Faryad Darabi Sahneh, Caterina Scoglio, and Piet Van Mieghem. Generalized epidemic mean-field model for spreading processes over multilayer complex networks. *IEEE/ACM Transactions on Networking*, 21(5):1609–1620, 2013.

[29] Faryad Darabi Sahneh, Caterina Scoglio, and Piet Van Mieghem. Generalized epidemic mean-field model for spreading processes over multilayer complex networks. *IEEE/ACM Transactions on Networking*, 21(5):1609–1620, 2013.

[30] Faryad Darabi Sahneh, Aram Vajdi, Heman Shakeri, Futing Fan, and Caterina Scoglio. Gemfsim: A stochastic simulator for the generalized epidemic modeling framework. *Journal of computational science*, 22:36–44, 2017.

[31] Daniel T Gillespie. Exact stochastic simulation of coupled chemical reactions. *The journal of physical chemistry*, 81(25):2340–2361, 1977.

[32] Wan Yang, Alicia Karspeck, and Jeffrey Shaman. Comparison of filtering methods for the modeling and retrospective forecasting of influenza epidemics. *PLoS Comput Biol*, 10(4):e1003583, 2014.

[33] Nedialko B Dimitrov and Lauren Ancel Meyers. Mathematical approaches to infectious disease prediction and control. In *Risk and Optimization in an Uncertain World*, pages 1–25. INFORMS, 2010.

[34] M. Sanjeev Arulampalam, Simon Maskell, Neil Gordon, and Tim Clapp. A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking. *IEEE Transactions on signal processing*, 50(2):174–188, 2002.

[35] HM Yang, M d L d G Macoris, KC Galvani, MTM Andrighetti, and DMV Wanderley. Assessing the effects of temperature on the population of aedes aegypti, the vector of dengue. *Epidemiology & Infection*, 137(8):1188–1202, 2009.

[36] Musa Sekamatte, Mahbubul H Riad, Tesfaalem Tekleghiorghis, Kenneth J Linthicum, Seth C Britch, Juergen A Richt, JP Gonzalez, and Caterina M Scoglio. Individual-based network model for rift valley fever in kabale district, uganda. *PloS one*, 14(3): e0202721, 2019.

[37] Influenza (seasonal), 2018. http://www.who.int/news-room/fact-sheets/detail/influenza-(seasonal).

[38] Mark O Nanyingi, Peninah Munyua, Stephen G Kiama, Gerald M Muchemi, Samuel M Thumbi, Austine O Bitek, Bernard Bett, Reese M Muriithi, and M Kariuki Njenga. A systematic review of rift valley fever epidemiology 1931–2014. *Infection ecology & epidemiology*, 5(1):28024, 2015.

[39] Natsuko Imai, Ilaria Dorigatti, Simon Cauchemez, and Neil M Ferguson. Estimating dengue transmission intensity from sero-prevalence surveys in multiple countries. *PLoS neglected tropical diseases*, 9(4):e0003719, 2015.

[40] Christopher W Woods, Adam M Karpati, Thomas Grein, Noel McCarthy, Peter Gaturuku, Eric Muchiri, Lee Dunster, Alden Henderson, Ali S Khan, Robert Swanepoel, et al. An outbreak of rift valley fever in northeastern kenya, 1997-98. *Emerging infectious diseases*, 8(2):138, 2002.

[41] Christine C Jost, Serge Nzietchueng, Simon Kihu, Bernard Bett, George Njogu, Emmanuel S Swai, and Jeffrey C Mariner. Epidemiological assessment of the rift valley fever outbreak in kenya and tanzania in 2006 and 2007. *The American journal of tropical medicine and hygiene*, 83(2_Suppl):65–72, 2010.

[42] Adel I Al-Afaleq and Mansour F Hussein. The status of rift valley fever in animals in saudi arabia: a mini review. *Vector-Borne and Zoonotic Diseases*, 11(12):1513–1520, 2011.

[43] Shaïf ABDO-SALEM, Guillaume Gerbier, Pascal Bonnet, Mohamed AL-QADASI, Annelise Tran, Etienne Thiry, G Al-Eryni, and F Roger. Descriptive and spatial epidemiology of rift valley fever outbreak in yemen 2000–2001. *Annals of the New York Academy of Sciences*, 1081(1):240–242, 2006.

[44] Osama Ahmed Hassan, Clas Ahlm, Rosemary Sang, and Magnus Evander. The 2007 rift valley fever outbreak in sudan. *PLoS neglected tropical diseases*, 5(9), 2011.

[45] S Jäckel, M Eiden, BO El Mamy, K Isselmou, A Vina-Rodriguez, B Doumbia, and MH Groschup. Molecular and serological studies on the r ift v alley fever outbreak in m auritania in 2010. *Transboundary and emerging diseases*, 60:31–39, 2013.

[46] Abdourahmane Sow, Ousmane Faye, Yamar Ba, Hampathé Ba, Diawo Diallo, Oumar Faye, Cheikh Loucoubar, Mohamed Boushab, Yahya Barry, Mawlouth Diallo, et al. Rift valley fever outbreak, southern mauritania, 2012. *Emerging infectious diseases*, 20(2):296, 2014.

[47] Osama Ahmed Hassan, Hippolyte Affognon, Joacim Rocklöv, Peter Mburu, Rosemary Sang, Clas Ahlm, and Magnus Evander. The one health approach to identify knowledge, attitudes and practices that affect community involvement in the control of rift valley fever outbreaks. *PLoS neglected tropical diseases*, 11(2):e0005383, 2017.

[48] RE Shope, CJ Peters, and FG Davies. The spread of rift valley fever and approaches to its control. *Bulletin of the World Health Organization*, 60(3):299, 1982.

[49] Maria Baudin, Ammar M Jumaa, Huda JE Jomma, Mubarak S Karsany, Göran Bucht, Jonas Näslund, Clas Ahlm, Magnus Evander, and Nahla Mohamed. Association of rift valley fever virus infection with miscarriage in sudanese women: a cross-sectional study. *The Lancet Global Health*, 4(11):e864–e871, 2016.

[50] Marisa Peyre, Véronique Chevalier, Shaif Abdo-Salem, An Velthuis, Nicolas Antoine-Moussiaux, Etienne Thiry, and François Roger. A systematic scoping study of the socio-economic impact of rift valley fever: Research gaps and needs. *Zoonoses and public health*, 62(5):309–325, 2015.

[51] Robert F Breiman, M Kariuki Njenga, Sarah Cleaveland, SK Sharif, Murithi Mbabu, and Lonnie King. Lessons from the 2006–2007 rift valley fever outbreak in east africa: implications for prevention of emerging infectious diseases. *Future Virology*, 3(5), 2008.

[52] AlaaEddeen M Seufi and Fatma H Galal. Role of culex and anopheles mosquito species as potential vectors of rift valley fever virus in sudan outbreak, 2007. *BMC infectious Diseases*, 10(1):65, 2010.

[53] FG Davies, KJ Linthicum, and AD James. Rainfall and epizootic rift valley fever. *Bulletin of the World Health Organization*, 63(5):941, 1985.

[54] Assaf Anyamba, Jean-Paul Chretien, Jennifer Small, Compton J Tucker, Pierre B Formenty, Jason H Richardson, Seth C Britch, David C Schnabel, Ralph L Erickson, and Kenneth J Linthicum. Prediction of a rift valley fever outbreak. *Proceedings of the National Academy of Sciences*, 106(3):955–959, 2009.

[55] Kenneth J Linthicum, Seth C Britch, and Assaf Anyamba. Rift valley fever: an emerging mosquito-borne disease. *Annual review of entomology*, 61:395–415, 2016.

[56] Necibe Tuncer, Hayriye Gulbudak, Vincent L Cannataro, and Maia Martcheva. Structural and practical identifiability issues of immuno-epidemiological vector–host models with application to rift valley fever. *Bulletin of mathematical biology*, 78(9):1796–1827, 2016.

[57] Uganda Bureau of Statistics – UBOS. Uganda bureau of statistics – ubos, 2018. https://www.ubos.org/.

[58] Saul C Mpeshe, Heikki Haario, and Jean M Tchuenche. A mathematical model of rift valley fever with human host. *Acta biotheoretica*, 59(3-4):231–250, 2011.

[59] David W Redding, Sonia Tiedt, Gianni Lo Iacono, Bernard Bett, and Kate E Jones. Spatial, seasonal and climatic predictive models of rift valley fever disease across africa. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 372 (1725):20160165, 2017.

[60] Alfred O Ochieng, Mark Nanyingi, Edwin Kipruto, Isabella M Ondiba, Fred A Amimo, Christopher Oludhe, Daniel O Olago, Isaac K Nyamongo, and Benson BA Estambale. Ecological niche modelling of rift valley fever virus vectors in baringo, kenya. *Infection ecology & epidemiology*, 6(1):32322, 2016.

[61] Mahbubul H Riad, Caterina M Scoglio, D Scott McVey, and Lee W Cohnstaedt. An individual-level network model for a hypothetical outbreak of japanese encephalitis in the usa. *Stochastic environmental research and risk assessment*, 31(2):353–367, 2017.

[62] Mathieu Bastian, Sebastien Heymann, and Mathieu Jacomy. Gephi: an open source software for exploring and manipulating networks. In *Third international AAAI conference on weblogs and social media*, 2009.

[63] Ernest A Gould and Stephen Higgs. Impact of climate change and other factors on emerging arbovirus diseases. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, 103(2):109–121, 2009.

[64] Mahbubul H Riad, Caterina M Scoglio, D Scott McVey, and Lee W Cohnstaedt. Estimation of parameters and basic reproductive ratio for japanese encephalitis transmission in the philippines using a sequential monte carlo filter. In *Control Technology and Applications (CCTA), 2017 IEEE Conference on*, pages 668–673. IEEE, 2017.

[65] Mahbubul H. Riad, Caterina M. Scoglio, D. Scott McVey, and Lee W. Cohnstaedt. An individual-level network model for a hypothetical outbreak of Japanese encephalitis in the USA. *Stochastic Environmental Research and Risk Assessment*, pages 1–15, 2016.

[66] Musa Sekamatte, Mahbubul H Riad, Tesfaalem Tekleghiorghis, Kenneth J Linthicum, Seth C Britch, Juergen A Richt, JP Gonzalez, and Caterina M Scoglio. Individual-based network model for rift valley fever in kabale district, uganda, to guide mitigation measures: A one health model. *bioRxiv*, page 388785, 2018.

[67] Cameron Nowzari, Victor M Preciado, and George J Pappas. Analysis and control of epidemics: A survey of spreading processes on complex networks. *IEEE Control Systems*, 36(1):26–46, 2016.

[68] BB Mukhopadhyay and PK Tapaswi. An sirs epidemic model of japanese encephalitis. *International Journal of Mathematics and Mathematical Sciences*, 17(2):347–355, 1994.

[69] Ram Naresh and Surabhi Pandey. Modeling and analysis of the spread of japanese encephalitis with environmental effects. *Appl. Appl. Math*, 4(1):155–175, 2009.

[70] Daihai He, Edward L Ionides, and Aaron A King. Plug-and-play inference for disease dynamics: measles in large and small populations as a case study. *Journal of the Royal Society Interface*, 2009.

[71] Narges Montazeri Shahtori, Caterina Scoglio, Arash Pourhabib, and Faryad Darabi Sahneh. Sequential monte carlo filtering estimation of ebola progression in west africa. In *American Control Conference (ACC), 2016*, pages 1277–1282. IEEE, 2016.

[72] Daniel M. Sheinson, Jarad Niemi, and Wendy Meiring. Comparison of the performance of particle filter algorithms applied to tracking of a disease epidemic. *Mathematical biosciences*, 255:21–32, 2014.

[73] Alex Skvortsov and Branko Ristic. Monitoring and prediction of an epidemic outbreak using syndromic observations. *Mathematical biosciences*, 240(1):12–19, 2012.

[74] Jane Liu and Mike West. Combined parameter and state estimation in simulation-based filtering. In *Sequential Monte Carlo methods in practice*, pages 197–223. Springer, 2001.

[75] Anna Lena Lopez, Josephine G Aldaba, Vito G Roque Jr, Amado O Tandoc III, Ava Kristy Sy, Fe Esperanza Espino, Maricel DeQuiroz-Castro, Youngmee Jee, Maria Joyce Ducusin, and Kimberley K Fox. Epidemiology of japanese encephalitis in the philippines: a systematic review. *PLoS Negl Trop Dis*, 9(3):e0003630, 2015.

[76] Christian Plagemann, Dieter Fox, and Wolfram Burgard. Efficient Failure Detection on Mobile Robots Using Particle Filters with Gaussian Process Proposals. In *IJCAI*, pages 2185–2190, 2007.

[77] John S Mackenzie, Duane J Gubler, and Lyle R Petersen. Emerging flaviviruses: the spread and resurgence of japanese encephalitis, west nile and dengue viruses. *Nature medicine*, 10:S98–S109, 2004.

[78] Adnan Khan, Muhammad Hassan, and Mudassar Imran. Estimating the basic reproduction number for single-strain dengue fever epidemics. *Infectious diseases of poverty*, 3(1):12, 2014.

[79] Clara Champagne, David Georges Salthouse, Richard Paul, Van-Mai Cao-Lormeau, Benjamin Roche, and Bernard Cazelles. Structure in the variability of the basic reproductive number (r0) for zika epidemics in the pacific islands. *eLife*, 5:e19874, 2016.

[80] P. Calistri, L. Savini, L. Candeloro, D. Di Sabatino, F. Cito, R. Bruno, and M. L. Danzetta. A Transitional Model for the Evaluation of West Nile Virus Transmission in Italy. *Transbound Emerg Dis*, 63(5):485–496, October 2016. ISSN 1865-1682. doi: 10.1111/tbed.12290.

[81] Mahbubul H Riad, Caterina M Scoglio, Lee W Cohnstaedt, and D Scott McVey. Short-term forecast and dual state-parameter estimation for japanese encephalitis transmis-

sion using ensemble kalman filter. In *2019 American Control Conference (ACC)*, pages 3444–3449. IEEE, 2019.

[82] Hamid Moradkhani, Soroosh Sorooshian, Hoshin V Gupta, and Paul R Houser. Dual state–parameter estimation of hydrological models using ensemble kalman filter. *Advances in water resources*, 28(2):135–147, 2005.

[83] Rudolph Emil Kalman et al. A new approach to linear filtering and prediction problems. *Journal of basic Engineering*, 82(1):35–45, 1960.

[84] Jeffrey Shaman and Alicia Karspeck. Forecasting seasonal outbreaks of influenza. *Proceedings of the National Academy of Sciences*, 109(50):20425–20430, 2012.

[85] Jeffrey Shaman, Alicia Karspeck, Wan Yang, James Tamerius, and Marc Lipsitch. Real-time influenza forecasts during the 2012–2013 season. *Nature communications*, 4, 2013.

[86] Geir Evensen. Sequential data assimilation with a nonlinear quasi-geostrophic model using monte carlo methods to forecast error statistics. *Journal of Geophysical Research: Oceans*, 99(C5):10143–10162, 1994.

[87] Gerrit Burgers, Peter Jan van Leeuwen, and Geir Evensen. Analysis scheme in the ensemble kalman filter. *Monthly weather review*, 126(6):1719–1724, 1998.

[88] Steven Gillijns, O. Barrero Mendoza, J. Chandrasekar, B. L. R. De Moor, D. S. Bernstein, and A. Ridley. What is the ensemble Kalman filter and how well does it work? In *American Control Conference, 2006*, pages 6–pp. IEEE, 2006.

[89] Kara E Rudolph, Justin Lessler, Rachael M Moloney, Brittany Kmush, and Derek AT Cummings. Incubation periods of mosquito-borne viral infections: a systematic review. *The American journal of tropical medicine and hygiene*, 90(5):882–891, 2014.

[90] Mahbubul H Riad, Musa Sekamatte, Felix Ocom, Issa Makumbi, and Caterina M

Scoglio. Risk assessment of ebola virus disease spreading in uganda using a two-layer temporal network. *Scientific reports*, 9(1):1–17, 2019.

[91] Marco Ajelli, Bruno Gonçalves, Duygu Balcan, Vittoria Colizza, Hao Hu, José J Ramasco, Stefano Merler, and Alessandro Vespignani. Comparing large-scale computational approaches to epidemic modeling: agent-based versus structured metapopulation models. *BMC infectious diseases*, 10(1):190, 2010.

[92] John J Potterat, SQ Muth, RB Rothenberg, H Zimmerman-Rogers, DL Green, JE Taylor, MS Bonney, and HA White. Sexual network structure as an indicator of epidemic phase. *Sexually transmitted infections*, 78(suppl 1):i152–i158, 2002.

[93] Alessandro Rizzo, Biagio Pedalino, and Maurizio Porfiri. A network model for ebola spreading. *Journal of theoretical biology*, 394:212–222, 2016.

[94] Nicola Perra, Bruno Gonçalves, Romualdo Pastor-Satorras, and Alessandro Vespignani. Activity driven modeling of time varying networks. *Scientific reports*, 2:469, 2012.

[95] Mark Newman, Albert-Laszlo Barabasi, and Duncan J Watts. *The structure and dynamics of networks*, volume 12. Princeton University Press, 2011.

[96] Christian L Althaus. Estimating the reproduction number of ebola virus (ebov) during the 2014 outbreak in west africa. *PLoS currents*, 6, 2014.

[97] Christian L Althaus, Nicola Low, Emmanuel O Musa, Faisal Shuaib, and Sandro Gsteiger. Ebola virus disease outbreak in nigeria: transmission dynamics and rapid control. *Epidemics*, 11:80–84, 2015.

[98] Maria Vittoria Barbarossa, Attila Dénes, Gábor Kiss, Yukihiko Nakata, Gergely Röst, and Zsolt Vizi. Transmission dynamics and final epidemic size of ebola virus disease outbreaks with varying interventions. *PloS one*, 10(7):e0131398, 2015.

[99] Adnan Khan, Mahim Naveed, Muhammad Dur-e Ahmad, and Mudassar Imran. Estimating the basic reproductive ratio for the ebola outbreak in liberia and sierra leone. *Infectious diseases of poverty*, 4(1):13, 2015.

[100] Marcelo FC Gomes, Ana Pastore y Piontti, Luca Rossi, Dennis Chao, Ira Longini, M Elizabeth Halloran, and Alessandro Vespignani. Assessing the international spreading risk associated with the 2014 west african ebola outbreak. *PLoS currents*, 6, 2014.

[101] Abhishek Pandey, Katherine E Atkins, Jan Medlock, Natasha Wenzel, Jeffrey P Townsend, James E Childs, Tolbert G Nyenswah, Martial L Ndeffo-Mbah, and Alison P Galvani. Strategies for containing ebola in west africa. *Science*, 346(6212): 991–995, 2014.

[102] Ying-Hen Hsieh, Junli Liu, Yun-Huei Tzeng, and Jianhong Wu. Impact of visitors and hospital staff on nosocomial transmission and spread to community. *Journal of theoretical biology*, 356:20–29, 2014.

[103] Marco Ajelli, Stefano Parlamento, David Bome, Atiba Kebbi, Andrea Atzori, Clara Frasson, Giovanni Putoto, Dante Carraro, and Stefano Merler. The 2014 ebola virus disease outbreak in pujehun, sierra leone: epidemiology and impact of interventions. *BMC medicine*, 13(1):281, 2015.

[104] Larry L Peterson and Bruce S Davie. *Computer networks: a systems approach*. Elsevier, 2007.

[105] Giuseppina La Rosa, Marta Fratini, Simonetta Della Libera, Marcello Iaconelli, and Michele Muscillo. Viral infections acquired indoors through airborne, droplet or contact transmission. *Annali dell'Istituto superiore di sanita*, 49:124–132, 2013.

[106] Juliette Stehlé, Nicolas Voirin, Alain Barrat, Ciro Cattuto, Vittoria Colizza, Lorenzo Isella, Corinne Régis, Jean-François Pinton, Nagham Khanafer, Wouter Van den Broeck, et al. Simulation of an seir infectious disease model on the dynamic contact network of conference attendees. *BMC medicine*, 9(1):87, 2011.

[107] Sebastián Risau-Gusmán and Damián H Zanette. Contact switching as a control strategy for epidemic outbreaks. *Journal of theoretical biology*, 257(1):52–60, 2009.

[108] Suyu Liu, Nicola Perra, Márton Karsai, and Alessandro Vespignani. Controlling contagion processes in activity driven networks. *Physical review letters*, 112(11):118702, 2014.

[109] Matthieu Nadini, Kaiyuan Sun, Enrico Ubaldi, Michele Starnini, Alessandro Rizzo, and Nicola Perra. Epidemic spreading in modular time-varying networks. *Scientific reports*, 8(1):2352, 2018.

[110] Alessandro Rizzo, Mattia Frasca, and Maurizio Porfiri. Effect of individual behavior on epidemic spreading in activity-driven networks. *Physical Review E*, 90(4):042801, 2014.

[111] Chen Liu, Li-xin Zhou, Chong-jun Fan, Liang-an Huo, and Zhan-wei Tian. Activity of nodes reshapes the critical threshold of spreading dynamics in complex networks. *Physica A: Statistical Mechanics and its Applications*, 432:269–278, 2015.

[112] Christian Bongiorno, Alessandro Rizzo, and Maurizio Porfiri. An information-theoretic approach to study activity driven networks. In *2018 IEEE International Symposium on Circuits and Systems (ISCAS)*, pages 1–5. IEEE, 2018.

[113] Christian Bongiorno, Lorenzo Zino, and Alessandro Rizzo. A novel framework for community modeling and characterization in directed temporal networks. *Applied Network Science*, 4(1):10, 2019.

[114] Narges Montazeri Shahtori, Tanvir Ferdousi, Caterina Scoglio, and Faryad Darabi Sahneh. Quantifying the impact of early-stage contact tracing on controlling ebola diffusion. *Mathematical Biosciences & Engineering*, 15(5):1165–1180, 2018.

[115] Lorenzo Zino, Alessandro Rizzo, and Maurizio Porfiri. An analytical framework for the study of epidemic models on activity driven networks. *Journal of Complex Networks*, 5(6):924–952, 2017.

[116] Yanjun Lei, Xin Jiang, Quantong Guo, Yifang Ma, Meng Li, and Zhiming Zheng. Contagion processes on the static and activity-driven coupling networks. *Physical Review E*, 93(3):032308, 2016.

[117] Matthieu Nadini, Alessandro Rizzo, and Maurizio Porfiri. Epidemic spreading in temporal and adaptive networks with static backbone. *IEEE Transactions on Network Science and Engineering*, 2018.

[118] Sifat A Moon, Lee W Cohnstaedt, D Scott McVey, and Caterina M Scoglio. A spatio-temporal individual-based network framework for west nile virus in the usa: spreading pattern of west nile virus. *PLoS computational biology*, 15(3):e1006875, 2019.

[119] Tanvir Ferdousi, Lee W Cohnstaedt, DS McVey, and Caterina M Scoglio. Understanding the survival of zika virus in a vector interconnected sexual contact network. *bioRxiv*, page 518613, 2019.

[120] Christian L Vestergaard and Mathieu Génois. Temporal gillespie algorithm: Fast simulation of contagion processes on time-varying networks. *PLoS computational biology*, 11(10):e1004579, 2015.

[121] Uganda population, 2019. http://worldpopulationreview.com/countries/uganda-population/.

[122] Douglas T Fleming and Judith N Wasserheit. From epidemiological synergy to public health policy and practice: the contribution of other sexually transmitted diseases to sexual transmission of hiv infection. *Sexually transmitted infections*, 75(1):3–17, 1999.

[123] WHO Ebola Response Team. Ebola virus disease in west africa—the first 9 months of the epidemic and forward projections. *New England Journal of Medicine*, 371(16): 1481–1495, 2014.

[124] Richard P Wenzel and Barbara B Carlson. Hospital epidemiology beyond infection control and toward quality assurance, 1988.

[125] Bing Wang, Lang Cao, Hideyuki Suzuki, and Kazuyuki Aihara. Safety-information-driven human mobility patterns with metapopulation epidemic dynamics. *Scientific reports*, 2:887, 2012.

[126] Shantesh Kumar Singh. Infectious diseases, national security and globalisation. *World Affairs: The Journal of International Issues*, 23(1):10–23, 2019.

[127] Susanne Rauch, Edith Jasny, Kim E Schmidt, and Benjamin Petsch. New vaccine technologies to combat outbreak situations. *Frontiers in immunology*, 9, 2018.

[128] Frederick G Hayden, Martin Friede, and Daniel G Bausch. Experimental therapies for ebola virus disease: what have we learned? *The Journal of infectious diseases*, 215(2): 167–170, 2017.

[129] L Sedda, DW Morley, MAH Braks, L De Simone, D Benz, and DJ Rogers. Risk assessment of vector-borne diseases for public health governance. *Public health*, 128 (12):1049–1058, 2014.

[130] Kimberly M Thompson, Paul F Deisler Jr, and Richard C Schwing. Interdisciplinary vision: the first 25 years of the society for risk analysis (sra), 1980–2005. *Risk Analysis: An International Journal*, 25(6):1333–1386, 2005.

[131] John S Brownstein, Hilary Rosen, Dianne Purdy, James R Miller, Mario Merlino, Farzad Mostashari, and Durland Fish. Spatial analysis of west nile virus: rapid risk assessment of an introduced vector-borne zoonosis. *Vector Borne and Zoonotic Diseases*, 2(3):157–164, 2002.

[132] R Romi, D Boccolini, M Menegon, and G Rezza. Probable autochthonous introduced malaria cases in italy in 2009–2011 and the risk of local vector-borne transmission. *Eurosurveillance*, 17(48):20325, 2012.

[133] CA Robin, Giovanni Lo Iacono, S Gubbins, JLN Wood, and JR Newton. The accuracy of the n ational e quine d atabase in relation to vector-borne disease risk modelling of horses in g reat b ritain. *Equine veterinary journal*, 45(3):302–308, 2013.

[134] Xinning Gui, Yubo Kou, Kathleen Pine, Elisa Ladaw, Harold Kim, Eli Suzuki-Gill, and Yunan Chen. Multidimensional risk communication: public discourse on risks during an emerging epidemic. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, page 214. ACM, 2018.

[135] Jolyon M Medlock and Steve A Leach. Effect of climate change on vector-borne disease risk in the uk. *The Lancet Infectious Diseases*, 15(6):721–730, 2015.

[136] Jolyon M Medlock, David Avenell, Iain Barrass, and Steve Leach. Analysis of the potential for survival and seasonal activity of aedes albopictus (diptera: Culicidae) in the united kingdom. *Journal of Vector Ecology*, 31(2):292–305, 2006.

[137] Cyril Caminade, Sari Kovats, Joacim Rocklov, Adrian M Tompkins, Andrew P Morse, Felipe J Colón-González, Hans Stenlund, Pim Martens, and Simon J Lloyd. Impact of climate change on global malaria distribution. *Proceedings of the National Academy of Sciences*, 111(9):3286–3291, 2014.

[138] Dominik Fischer, Stephanie Margarete Thomas, Franziska Niemitz, Björn Reineking, and Carl Beierkuhnlein. Projection of climatic suitability for aedes albopictus skuse (culicidae) in europe under climate change conditions. *Global and Planetary Change*, 78(1-2):54–64, 2011.

[139] Dominik Fischer, Stephanie M Thomas, Jonathan E Suk, Bertrand Sudre, Andrea Hess, Nils B Tjaden, Carl Beierkuhnlein, and Jan C Semenza. Climate change effects on chikungunya transmission in europe: geospatial analysis of vector's climatic suitability and virus' temperature requirements. *International journal of health geographics*, 12 (1):51, 2013.

[140] MW Service. Ecological and biological studies on aedes cantans (meig.)(diptera: Culicidae) in southern england. *Journal of Applied Ecology*, pages 159–196, 1977.

[141] DJ Rogers and SE Randolph. Climate change and vector-borne diseases. *Advances in parasitology*, 62:345–381, 2006.

[142] Ramesh C Dhiman, Sharmila Pahwa, GPS Dhillon, and Aditya P Dash. Climate change and threat of vector-borne diseases in india: are we prepared? *Parasitology research*, 106(4):763–773, 2010.

[143] P Gale, A Brouwer, V Ramnial, L Kelly, R Kosmider, AR Fooks, and EL Snary. Assessing the impact of climate change on vector-borne viruses in the eu through the elicitation of expert opinion. *Epidemiology & Infection*, 138(2):214–225, 2010.

[144] C Garrett-Jones. The human blood index of malaria vectors in relation to epidemiological assessment. *Bulletin of the World Health Organization*, 30(2):241, 1964.

[145] C Garrett-Jones and GR Shidrawi. Malaria vectorial capacity of a population of anopheles gambiae: an exercise in epidemiological entomology. *Bulletin of the World Health Organization*, 40(4):531, 1969.

[146] Jing Liu-Helmersson, Hans Stenlund, Annelies Wilder-Smith, and Joacim Rocklöv. Vectorial capacity of aedes aegypti: effects of temperature and implications for global dengue epidemic potential. *PloS one*, 9(3):e89783, 2014.

[147] Thomas W Scott, Priyanie H Amerasinghe, Amy C Morrison, Leslie H Lorenz, Gary G Clark, Daniel Strickman, Pattamaporn Kittayapong, and John D Edman. Longitudinal studies of aedes aegypti (diptera: Culicidae) in thailand and puerto rico: blood feeding frequency. *Journal of medical entomology*, 37(1):89–101, 2000.

[148] Jean-Francois Briere, Pascale Pracros, Alain-Yves Le Roux, and Jean-Sebastien Pierre. A novel rate model of temperature-dependent development for arthropods. *Environmental Entomology*, 28(1):22–29, 1999.

[149] Louis Lambrechts, Krijn P Paaijmans, Thanyalak Fansiri, Lauren B Carrington, Laura D Kramer, Matthew B Thomas, and Thomas W Scott. Impact of daily temperature fluctuations on dengue virus transmission by aedes aegypti. *Proceedings of the National Academy of Sciences*, 108(18):7460–7465, 2011.

[150] Dana A Focks, Eric Daniels, Dan G Haile, and James E Keesling. A simulation model of the epidemiology of urban dengue fever: literature analysis, model development, preliminary validation, and samples of simulation results. *The American journal of tropical medicine and hygiene*, 53(5):489–506, 1995.

[151] Joceline Lega, Heidi E Brown, and Roberto Barrera. Aedes aegypti (diptera: Culicidae) abundance model improved with relative humidity and precipitation-driven egg hatching. *Journal of medical entomology*, 54(5):1375–1384, 2017.

[152] Marian Boguná, Luis F Lafuerza, Raúl Toral, and M Ángeles Serrano. Simulating non-markovian stochastic processes. *Physical Review E*, 90(4):042108, 2014.

[153] Naoki Masuda and Luis EC Rocha. A gillespie algorithm for non-markovian stochastic processes. *SIAM Review*, 60(1):95–115, 2018.

[154] Lyle R Petersen, Kay M Tomashek, and Brad J Biggerstaff. Estimated prevalence of dengue viremia in puerto rican blood donations, 1995 through 2010. *Transfusion*, 52 (8):1647–1651, 2012.

[155] Henrik Salje, Justin Lessler, Kishor Kumar Paul, Andrew S Azman, M Waliur Rahman, Mahmudur Rahman, Derek Cummings, Emily S Gurley, and Simon Cauchemez. How social structures, space, and behaviors shape the spread of infectious diseases using chikungunya as a case study. *Proceedings of the National Academy of Sciences*, 113 (47):13420–13425, 2016.

[156] Isabel Rodríguez-Barraquer, Rome Buathong, Sopon Iamsirithaworn, Ananda Nisalak, Justin Lessler, Richard G Jarman, Robert V Gibbons, and Derek AT Cummings. Revisiting rayong: shifting seroprofiles of dengue in thailand and their implications for transmission and control. *American journal of epidemiology*, 179(3):353–360, 2013.

[157] Isabel Rodríguez-Barraquer, Sunil S Solomon, Periaswamy Kuganantham, Aylur Kailasom Srikrishnan, Canjeevaram K Vasudevan, Syed H Iqbal, Pachamuthu Balakrishnan, Suniti Solomon, Shruti H Mehta, and Derek AT Cummings. The hidden

burden of dengue and chikungunya in chennai, india. *PLoS neglected tropical diseases*, 9(7):e0003906, 2015.

[158] Mahbubur Rahman, Khalilur Rahman, AK Siddque, Shereen Shoma, AHM Kamal, KS Ali, Ananda Nisaluk, and Robert F Breiman. First outbreak of dengue hemorrhagic fever, bangladesh. *Emerging infectious diseases*, 8(7):738, 2002.

[159] Sifat Sharmin, Elvina Viennet, Kathryn Glass, and David Harley. The emergence of dengue in bangladesh: epidemiology, challenges and future disease risk. *Transactions of The Royal Society of Tropical Medicine and Hygiene*, 109(10):619–627, 2015.

[160] Emran Bin Yunus, A Mannan Bangali, M Mahmood, M Mushfiqur Rahman, AR Chowdhury, and KR Talukder. Dengue outbreak 2000 in bangladesh: From speculation to reality and exercises. *Dengue Bulletin*, 25, 2001.

[161] S Bhatt, PW Gething, OJ Brady, JP Messina, AW Farlow, CL Moyes, JM Drake, JS Brownstein, AG Hoen, O Sankoh, et al. The global distribution and burden of dengue. nature496: 504–507, 2013.

[162] Eduardo A Undurraga, Yara A Halasa, and Donald S Shepard. Use of expansion factors to estimate the burden of dengue in southeast asia: a systematic analysis. *PLoS neglected tropical diseases*, 7(2):e2056, 2013.

[163] Parnali Dhar-Chowdhury, Kishor Kumar Paul, C Emdad Haque, Shakhawat Hossain, L Robbin Lindsay, Antonia Dibernardo, W Abdullah Brooks, and Michael A Drebot. Dengue seroprevalence, seroconversion and risk factors in dhaka, bangladesh. *PLoS neglected tropical diseases*, 11(3):e0005475, 2017.

[164] Miranda Chan and Michael A Johansson. The incubation periods of dengue viruses. *PloS one*, 7(11), 2012.

[165] Climate data for cities worldwide, 2020. https://en.climate-data.org.

[166] City population, 2020. https://www.citypopulation.de/.

[167] Sifat Sharmin, Kathryn Glass, Elvina Viennet, and David Harley. Interaction of mean temperature and daily fluctuation influences dengue incidence in dhaka, bangladesh. *PLoS neglected tropical diseases*, 9(7), 2015.

[168] Chee-Fu Yung, Kim-Sung Lee, Tun-Linn Thein, Li-Kiang Tan, Victor C Gan, Joshua GX Wong, David C Lye, Lee-Ching Ng, and Yee-Sin Leo. Dengue serotype-specific differences in clinical manifestation, laboratory parameters and risk of severe disease in adults, singapore. *The American journal of tropical medicine and hygiene*, 92(5):999–1005, 2015.

[169] Arifa Akram. Alarming turn of dengue fever in dhaka city in 2019. *Bangladesh Journal of Infectious Diseases*, 6(1):1–2, 2019.

[170] T Shirin, AKM Muraduzzaman, AN Alam, S Sultana, M Siddiqua, MH Khan, A Akram, AR Sharif, S Hossain, and MS Flora. Largest dengue outbreak of the decade with high fatality may be due to reemergence of den-3 serotype in dhaka, bangladesh, necessitating immediate public health attention. *New microbes and new infections*, 29, 2019.

[171] AKM Muraduzzaman, Ahmed Nawsher Alam, Sharmin Sultana, Mahmuda Siddiqua, Manjur Hossain Khan, Arifa Akram, Farhana Haque, Meerjady Sabrina Flora, and Tahmina Shirin. Circulating dengue virus serotypes in bangladesh from 2013 to 2016. *Virusdisease*, 29(3):303–307, 2018.

[172] Md Mahbubul Huq Riad. Modeling japanese encephalitis using interconnected networks for a hypothetical outbreak in the usa. *KREX*, 2017.

[173] RS Lanciotti, JT Roehrig, V Deubel, J Smith, M Parker, K Steele, B Crise, KE Volpe, MB Crabtree, JH Scherret, et al. Origin of the west nile virus responsible for an outbreak of encephalitis in the northeastern united states. *Science*, 286(5448):2333–2337, 1999.

[174] O. L. Weaver. Kansas State University, Manhattan, KS, 2005. (Private Communication).

[175] John H Rappole, Bradley W Compton, Peter Leimgruber, Jamie Robertson, David I King, and Swen C Renner. Modeling movement of west nile virus in the western hemisphere. *Vector-borne & zoonotic diseases*, 6(2):128–139, 2006.

[176] A Townsend Peterson, David A Vieglais, and James K Andreasen. Migratory birds modeled as critical transport agents for west nile virus in north america. *Vector-Borne and Zoonotic Diseases*, 3(1):27–37, 2003.

[177] M Agarwal and V Verma. The impact of media on the spreading and control of japanese encephalitis. *Int J Math Sci Comput*, 2(2):23–31, 2012.

[178] Aristides Moustakas and Matthew R Evans. Coupling models of cattle and farms with models of badgers for predicting the dynamics of bovine tuberculosis (tb). *Stochastic Environmental Research and Risk Assessment*, 29(3):623–635, 2015.

[179] Meret E Ricklin, Obdulio García-Nicolás, Daniel Brechbühl, Sylvie Python, Beatrice Zumkehr, Antoine Nougairede, Remi N Charrel, Horst Posthaus, Anna Oevermann, and Artur Summerfield. Vector-free transmission and persistence of japanese encephalitis virus in pigs. *Nature communications*, 7(1):1–9, 2016.

[180] West nile virus human provisional 2015 data, 2016. http://diseasemaps.usgs.gov/mapviewer/.

[181] Feral swine distribution map, 2016. http://swine.vet.uga.edu/nfsms/information/.

[182] Cornell lab of ornithology, 2016. http://ebird.org/ebird/places?yr=allandm.

[183] K Gunasekaran, SS Sahu, and P Jambulingam. Estimation of vectorial capacity of anopheles minimus theobald & an. fluviatilis james (diptera: Culicidae) in a malaria endemic area of odisha state, india. *The Indian journal of medical research*, 140(5): 653, 2014.

[184] Kevin Desouza and Lin Yuan. Towards evidence-driven policy design: complex adaptive systems and computational modeling. *Annual Review of Policy Design*, 1(1):1–19, 2013.

[185] G Wayne Brodland. How computational models can help unlock biological systems. In *Seminars in cell & developmental biology*, volume 47, pages 62–73. Elsevier, 2015.

[186] Andreas Handel, Ira M Longini Jr, and Rustom Antia. What is the best control strategy for multiple infectious disease outbreaks? *Proceedings of the Royal Society B: Biological Sciences*, 274(1611):833–837, 2007.

[187] Robert C Reiner Jr, T Alex Perkins, Christopher M Barker, Tianchan Niu, Luis Fernando Chaves, Alicia M Ellis, Dylan B George, Arnaud Le Menach, Juliet RC Pulliam, Donal Bisanzio, et al. A systematic review of mathematical models of mosquito-borne pathogen transmission: 1970–2010. *Journal of The Royal Society Interface*, 10(81): 20120921, 2013.

[188] Matthew R Evans, Tim G Benton, Volker Grimm, Catherine M Lessells, Maureen A O'Malley, Aristides Moustakas, and Michael Weisberg. Data availability and model complexity, generality, and utility: a reply to lonergan. *Trends in ecology & evolution*, 29(6):302–303, 2014.

[189] Jennifer Owen, Frank Moore, Nicholas Panella, Eric Edwards, Rachel Bru, Megan Hughes, and Nicholas Komar. Migrating birds as dispersal vehicles for west nile virus. *EcoHealth*, 3(2):79, 2006.

[190] Elsa Jourdain, Michel Gauthier-Clerc, Dominique Bicout, and Philippe Sabatier. Bird migration routes and risk for pathogen dispersion into western mediterranean wetlands. *Emerging infectious diseases*, 13(3):365, 2007.

[191] Kurt D Reed, Jennifer K Meece, James S Henkel, and Sanjay K Shukla. Birds, migration and emerging zoonoses: West nile virus, lyme disease, influenza a and enteropathogens. *Clinical medicine & research*, 1(1):5–12, 2003.

[192] Ian Newton. *The migration ecology of birds.* Elsevier, 2010.

Appendix A

An individual-level network model for a hypothetical outbreak of Japanese Encephalitis in the USA [1]

A.1 Introduction

The incidence of Japanese encephalitis (JE) has not been reported in the United States, but this pathogen belongs to the Japanese encephalitis group, the same group of the West Nile virus (WNV)[173]. Both these pathogens have a similar transmission cycle, which includes birds and mosquitoes[174]. The occurrence of WNV in the United States generated extensive research activities on its incidence data and transmission. Epidemiology of WNV and incidence data are useful for understanding the hypothetical introduction and incidence of JE in the United States. Therefore, careful observation of WNV epidemic models in the USA plays a major role in finding important factors in our JE model. Numerous nonlinear differential equation models —similar to those of JE— are available for WNV. Several models have been proposed based on the contribution of birds to the WNV transmission cycle. Rappole et al. modeled local and migratory birds spatially to understand their effects

[1] This chapter is a reformatted and slightly modified version of our published article[61;172]

of on WNV spread[175]. However, spreading pattern of WNV were not consistent with the flying pattern of migratory birds but local birds. Other models, however, showed that the migratory pattern of birds coincides with the spread of WNV in the United States[68;176;177], leading to positive and negative opinions about the importance of migration in the spread of WNV. Therefore, although some researchers deny that migratory birds contribute to the long-distance spread of pathogens, pathogens are likely to be transferred to new, distant locations via them. Aforementioned are important articles on JE and WNV modeling but may not provide comprehensive knowledge to determine which factors to include the model for JE in the US. Most of these models include human populations, which does not have an impact on the spreading process for being the dead-end hosts. Which is important because infectious disease models need to incorporate sufficient details about the modeled system to biologically and epidemiologically accurate outcomes[178]. Model parameters must still be precisely selected. However, accurate estimate of these parameters is challenging and prone to large errors in practice. If there are errors in these estimates, the prediction from the model about disease dynamic will be misleading. Therefore, the tradeoff between the number of parameters and the inclusion of minute details make the predictive modeling very challenging. In our individual-level network model, we limit the number of parameters. The literature on JE mainly focuses on Southeast Asia and Pacific where the virus circulates endemically as both a chronic risk in the south and outbreak hazard in the north. However, due to differences in mosquito vector species, and host species, an outbreak in the United States of America will likely have a different epidemiology. Our goal is to examine likely JE transmission along the north-south bird migration route on the east coast of US in the event of accidental or intentional introduction. Our disease model will elucidate the role and key interactions between various native reservoir populations (insect, avian, and mammalian) which may be involved in the pathogen transmission of this exotic virus in the US. Possible mitigation strategies will be tested to determine the most likely to reduce pathogen spread between geographic locations.

In this paper, we propose an individual-level network model of JE with probable US host and vector species in three spatially separated locations. Populations are selected for

our model based on extensive research about competent populations for JE in the US. We explicitly model pig populations and implicitly model mosquito and bird populations to examine JE epidemiology for our selected scenario. Though our proposed model is described for only one migration route but has the ability to be adapted for other routes by selecting alternate locations and host species on that route.

Our modeling approach is novel in its individual-level realization of feral pigs with mosquitoes and birds as transmission medium. Contact network among feral pigs within each location are created with two homogeneous network topologies- Fully connected and Erdos-Renyi. However, we use a heterogeneous contact structure among local feral pigs and distant feral pigs. This heterogeneity in the contact among local feral pigs is reflected using different pathogen transfer rates from infected to susceptible pigs. Therefore, well-known meta-population approach is not suitable for our network model. Heterogeneity in the contact structure necessitates an individual-level model to describe the epidemiology of JE. Our model being the individual-level, has the flexibility to incorporate the heterogeneity in the network topology when specific data is available about contact structure among individual feral pigs.

Our model results predict the maintenance of JE pathogen among birds once introduced via a migratory bird even in the absence of feral pigs which eventually results in human incidences. The number of humans infected with JE is comparable to WNV cases among human in our selected location. An effective mitigation strategy against JE is deduced from simulation results of our model. JE transmission can be reduced by lowering mosquito abundance, but both mosquitoes and birds need to be limited/controlled to prevent pathogen spread to distant locations especially in areas of high mosquito abundance.

A.2 Materials and method

A.2.1 The model

We develop a model for a scenario of JE epidemiology in the United States in which only one population—feral pigs in three spatial locations — is represented at the individual animal level via a connected network. Our simulation model uses a generalized epidemic modeling framework (GEMF)[29] developed by the Network Science and Engineering (NetSE) group at Kansas State University. We carry out extensive simulations, varying parameter values to determine their effects on the overall number of infections (total number of pigs infected) and to relate them to disease dynamics. Model scenarios that included fall and spring migration periods of birds are considered for each of following two network topologies: local fully connected and local Erdos-Renyi. In a local fully connected network, each feral pig is connected to all other pigs in the same location while in Erdos-Renyi network, connections between local feral pigs are random with a defined probability (Erdos and Renyi 1960). These topologies are used to create links among feral pigs within an individual location while inter-location links are created with a probability which ensured, at least, one link between them. In a fully connected network, links are created from a single feral pig to all other pigs in the same location. Therefore, if a location has K feral pigs, then it will have (K-1) links connecting it to all other feral pigs there. For Erdos-Renyi network, the connection between two feral pigs is created following three steps -

- from an infected pig to a susceptible pig via mosquito
- from an infected pig to a susceptible local bird and then from that infected bird to a susceptible pig
- (iii) from an infected pig to a susceptible migratory bird and then from that infected bird to a susceptible pig in a distant location

All transmissions occur by local competent mosquito blood feeding.

The role of mosquito populations in JE transmission to pigs is expressed by a parameter (β_1) which is the vectorial capacity of focal putative US mosquito vectors. Vectorial capacity β_1 is given as

$$V_c = \frac{ma^2 b_h b_m e^{-\mu_m n}}{\mu_m} \tag{A.1}$$

where the vector parameters used are 1) the average daily vector biting rate (a), 2) the probability of vector to human transmission per bite (b_h), 3) the probability of human to vector infection per bite (b_m), 4) the duration of the extrinsic incubation period (n), 5) the vector mortality rate (μ_m), and 6) mosquito vector density with respect to the host (m)[144–146]. These parameters are specific for the mosquito species and the concerned disease.

Infection rate that includes both mosquito and bird is expressed as β_2. Transfer rate β_2 is dependent on β_1 because this pathogen transfer also occurs via mosquitoes but the inclusion of birds made it different than vectorial capacity (β_1). Here, β_2 is expressed as rβ_1, where r is the local bird's contribution to pathogen spread with respect to the pig density. We refer r as bird community parameter which is a nonnegative parameter with a maximum value of 0.5 because transmission requires a minimum of four feedings when involving birds (pig –mosquito-bird–mosquito-pig) compared to two feedings without them (pig –mosquito -pig). When the bird population has an equal size to the feral pig population in a location, then the maximum value of r is possible. If the bird population exceeds the pig population, then the probability of an infected bird being bitten a second time decreases because alternate bird and pig hosts will be plentiful decreasing the probability of a second feeding (host saturation). Consequently, r decreases from its maximum value of 0.5 if the bird population is more or less than the pig one. We assumed this as we considering no host preference among mosquitoes due to lack of data about mosquito biting pattern among feral pigs and birds. However, our model being individual-level has the flexibility to include host preference when specific data is available. For our simulations, three different values of r for the bird community parameter are used: r=0.15 (low bird numbers and species diversity), 0.3 (medium numbers and species diversity) and 0.5 (high numbers and a diverse bird community). As three distant

geographic locations are considered in this model, viremic migratory birds are the only means of interstate pathogen transfer within our simulation period of a single migratory season (90-day). However, a number of conditions are necessary— a bird must become infected in the first location, travel to another location while remaining viremic and spread the infection to mosquitoes. However, no exact method is available to determine the probability of the occurrence of this event, but this is crucial for long distance spread of JE pathogen. This long distance pathogen transfer occurs with a rate β_3, which is dependent on β_2 of the beginning and ending or staging places of the migration and the number and diversity of migratory bird's species in the origin location. In summary, if there is link between two feral pigs, pathogen transmission between them is possible and this transmission occurs with a rate—

- β_1 if the link is between feral pigs within a location and pathogen transmission happens via (pig-mosquito-pig) cycle
- β_2 if the link is between feral pigs within a location and pathogen transmission happens via (pig-mosquito-bird-mosquito-pig) cycle
- β_3 if the link is between feral pigs in two distant locations

If there is no link between pigs (local or distant), the rate of pathogen transmission is zero. Whether the pathogen transfer from an infected to a susceptible local pig will happen with rate β_1 or β_2 is dependent on the respective bird community parameter. Links among local pigs are assigned a pathogen transfer rate β_2 with a homogeneous distribution of probability $\frac{r}{r+1}$. This ratio is assumed to reflect the relative abundance of birds with respect to feral pigs. Rest of the links are assigned a pathogen transfer rate β_1. However, pathogen transmissions occur only if a sufficient number of JE-competent mosquitoes bite the infected pig and, after an appropriate period of viral replication in the mosquito, feed on a susceptible pig. Infection processes are statistically independent, therefore, the transition rate for a susceptible node to the exposed state is the sum of transfer rates times the number of infected neighbor nodes. The total rate at which an individual pig can become infected is proportional to the number of infected pigs in the neighborhood and the population

size (or density) of competent mosquito vectors. The exposed compartment represents the delays for a susceptible individual to become infectious. An exposed node then become infectious with a rate, δ. This parameter value is dependent only on the time taken by the pigs to become infectious once they become exposed to JE pathogen. This parameter can have a value in the range 0.25 day^{-1}. It is kept constant at δ=0.4 day^{-1} (average) for all simulations in this paper as pigs take 1-4 days to become infectious once exposed to pathogen[179]. This constant value of is the mean of an exponential distribution used for this parameter during simulation with GEMF. However, a change in this parameter will change the speed of pathogen spreading from infected to susceptible pigs. Therefore, each individual pig can be in one of the three compartments—Susceptible, Exposed and Infectious. The transition of pigs from susceptible to exposed happens with rate β_1, β_2, β_3 and from exposed to infectious happens with incubation rate δ. For our model, we consider only one infected pig at the beginning. We did not consider any recovered compartment, as pigs remain infectious once infected. In the *Susceptible-Exposed-Infected (SEI)* model based on GEMF, infection processes are independent Poisson processes as shown in 1.1- **??**.

A.2.2 Network structure

Our network consists of three spatially separated locations: Miami-Dade County in Florida, Carteret County in North Carolina, and Charleston County in South Carolina. These three counties are selected because they provided WNV incidence data[180], have an abundance of feral pigs[181], the highest number of observed bird's species[182], and proximity to coastal areas. A simplified visualization of our model is presented in Figure A.1. Locations are selected from three states to encompass a wide range of variability in weather and habitat. Within each location, we consider a small geographical area of 60 sq mi or less as feral pigs roam approximately 10-60 sq mi in search of food[77]. Florida contains more feral pigs and ardeid birds than the other two locations. Birds and mosquitoes are highly variable from season to season and throughout the year, therefore, selection of the appropriate time frame to simulate is essential for accurate epidemiological estimations. Mosquitoes are abundant in the late

summer (June-July) and early fall (August-September), crucial seasons for the simulation as bird migration also occurs within this time period. Spring migration is northbound and takes place at the end of the spring and beginning of the summer (April-June). At this period, there is an abundance of birds migrating from the FL location to SC and NC locations but lower mosquito abundance as it occurs before late summer—summer being the peak time for mosquito abundance in all locations. Fall migration, however, occurs during late summer and early fall (July-September), coinciding with an abundant mosquito population. This time frame provides an abundance of mosquitoes, resident birds, and migrating birds in all three locations, which are crucial factors in long-distance dispersal and local transmission of JE.

The long-distance pathogen transmission is unidirectional—northbound during spring migration between three locations and southbound during fall. Within each geographical location, resident birds, mosquitoes, and pigs can move randomly in any direction. Pathogen transfer is highly dependent on the migration pattern of birds as various species have distinctive intervals between the staging places (places where birds take a break while migrating) and unique flying speeds while migrating.

A.2.3 Estimations and assumptions

Pig population data used for our simulations are derived from the feral pig mapping system[181] These published maps of the distribution and density of feral pigs throughout the United States allow us to determine the estimated number of feral pigs in various locations. We consider all three of our locations having a medium density of feral pigs— 10 animals per sq mi[182]. Consequently, 600 pigs in FL (60 sq mi area), 500 pigs in SC (50 sq mi) and 300 pigs in NC (30 sq mi) are selected. Therefore, our model contains a total of 1400 pigs (nodes in the network) within the three locations. We start each simulation with a single infected pig in the initial location of the bird migration. Therefore, we assume one infected pig in FL and NC location respectively for spring and fall migration. Selection of other density (low, high) of pigs would change the population in each location. Therefore we would have results

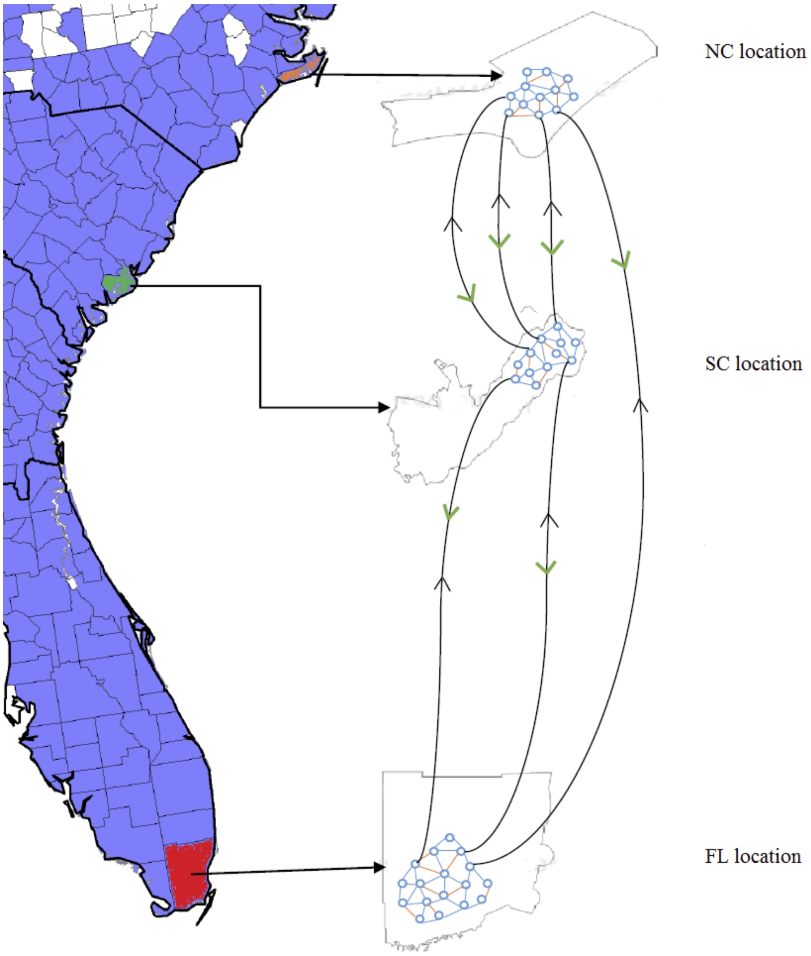


Figure A.1: *Purple shades in the map indicate the presence of feral pigs*[181] *Green arrows indicate directions of migration during fall migration period, and black arrows indicate the direction of migration during spring migration period. Blue circles represent feral pigs, blue lines represent direct links for possible transmission of the JE virus via mosquitoes, and orange lines represent links for possible transmission via mosquitoes and birds.*

with similar trends but different quantitative values.

In our model, we use a complex weighting system, a crucial factor in our simulation model, as weights are used to reflect the heterogeneity in mosquito and bird populations in different locations. Weights represent the temporal and spatial dependence of mosquito vectorial capacity β_1 in the different locations during fall and spring migrations. Mosquitoes are abundant in all three locations during fall migration, but the FL location always contains more mosquitoes than the other locations. Therefore, the selected weight ratio of FL: SC: NC during fall migration is 3:2:1 (values are 1.5 β_1, β_1, and 0.5 β_1 respectively) and 4:2:1 during

spring migration (values are β_1, 0.5 β_1 and 0.25 β_1 respectively). We chose mosquito vectorial capacity values for our simulations in such a way that they remain within a realistic range after the weighting[183]. Weights ratios reflect the relative abundance of mosquito within each simulation period (fall and spring) while the value of weights represents the actual abundance of mosquito. An important point here— this is just one scenario chosen for simulation purpose in this model, it could have been chosen otherwise by reflecting the higher abundance of mosquito during fall than spring migration period. In our result section, we express all Figures only through β_1 and r; However, different values of β_1 in each location are considered following the weighting system, as parameters are weighted with corresponding weights of that location and season.

The values of r for NC, SC, and FL are weighted for all simulations with $\frac{22}{22}$, $\frac{20}{23}$, and $\frac{29}{33}$, respectively. This set of weights are derived from a list of WNV- and JE-competent birds of 33 total species and the availability of bird species from that list in the corresponding locations. For example, the FL location contains 29 bird species out of the 33 competent species, hence the weight $\frac{29}{33}$. The pathogen transfer rates β_3 via migratory birds are expressed with the vectorial capacity of two locations. The migration process being independent of the origin and destination locations, β_3 is expressed as the product of β_2 of two locations and weighted with the fraction of the number of migratory species for each location to the total number of species. Therefore, the weights for β_3 for FL, SC and NC are respectively $\frac{18}{33}$, $\frac{14}{33}$ and $\frac{15}{33}$.

A.2.4 Mathematical model summary

A graph $G = \{V, E\}$ is considered for the population of feral pigs and their potential infectious dynamics among the three different spatial locations. The set of nodes V, with V = 1400, is given as $V = V_{FL} \cap V_{NC} \cap V_{SC}$ with $V_{FL} = 600$, $V_{NC} = 500$ and $V_{SC} = 300$. E is the set of links among individual feral pigs within and between locations. For any node $i \in V$, δ=0.4 $days^{-1}$ Each susceptible pig $i \in V$ can infect each susceptible feral pig $j \in V$ with an infectious rate $\beta_{i,j}$. Rate of pathogen transmission (incubation rate) among feral pigs

within the same location. We have two cycles of pathogen transfer among feral pigs within each location. They are— i) Pig –mosquito-pig cycle and ii) Pig-mosquito-bird-mosquito-pig cycle. We have assigned transfer rates to links among feral pigs with one of these two cycles. Distribution of transfer rates among local links is performed with a probability $\frac{r}{r+1}$ which is dependent on respective bird community parameter. The pathogen transfer rate via (pig –mosquito-pig) cycle is- Fall migration period $\beta^{int}_{i,j,fall}$= $1.5\beta_1$, $i, j \in FL$ $\beta^{int}_{i,j,fall}$= β_1, $i, j \in NC$ $\beta^{int}_{i,j,fall}$= $0.5\beta_1$, $i, j \in SC$

Spring migration period $\beta_{i,j}$.spring.int=β_1, $i, j \in FL$ $\beta^{ext}_{i,j,spring}$= $0.5\beta_1$, $i, j \in NC$ $\beta^{ext}_{i,j,spring}$= $0.25\beta_1$, $i, j \in SC$

The pathogen transfer rate via (pig –mosquito-bird-mosquito–pig) cycle is- Fall migration period

- $\beta^{int}_{i,j,fall}$= 1.5 $r_F L\beta_1$, $i, j \in FL$
- $\beta_{i,j}$.fallint.= $r_N C\beta_1$, $i, j \in NC$
- $\beta^{int}_{i,j,fall}$=0.5 $r_S C\beta_1$, $i, j \in SC$

Spring migration period

- $\beta_{i,j}$.spring.int= $r_F L\beta_1$, $i, j \in FL$
- $\beta^{ext}_{i,j,spring}$= 0.5 $r_N C\beta_1$, $i, j \in NC$
- $\beta^{ext}_{i,j,spring}$= 0.25 $r_S C\beta_1$, $i, j \in SC$

Where $r_F L = \frac{29}{33}r$, $r_N C = \frac{22}{33}r$, $r_S C = \frac{20}{33}r$ with r= bird community parameter. Rate of pathogen transmission among feral pigs within distant locations (β_3) are expressed as-

Fall migration period

- $\beta^{ext}_{i,j,fall}$= $\frac{15}{33}$*0.5*1.5* $r_N C r_F L$ $\beta_1 2$, $i \in NC$, $j \in FL$
- $\beta^{ext}_{i,j,fall}$= $\frac{15}{33}$*0.5 *1* $r_N C r_S C$ $\beta_1 2$, $i \in NC$, $j \in SC$
- $\beta^{ext}_{i,j,fall}$= $\frac{14}{33}$*1*1.5* $r_S C r_F L$ $\beta_1 2$, $i \in SC$, $j \in FL$

Spring migration period

- $\beta^{ext}_{i,j,spring}$= $\frac{18}{33}$*1*0.25* $r_F Lr_N C$ $\beta_1 2$, $i \in FL$, $j \in NC$
- $\beta^{ext}_{i,j,spring}$= $\frac{14}{33}$*0.5*0.25* $r_S Cr_N C$ $\beta_1 2$, $i \in SC$, $j \in NC$
- $\beta^{ext}_{i,j,spring}$.= $\frac{18}{33}$*1*0.5* $r_F Lr_S C$ $\beta_1 2$, $i \in FL$, $j \in SC$

Here $i \in$ origin location of migration and $j \in$ destination location of migration. Rates are weighted with the ratio of migratory species number to total 33 species. The infectious rates for other location pairs other than mentioned in the expression of β_3 are zero because bird migration is directional and is southbound during fall while northbound during spring migration.

Pathogens transfer from infected to susceptible pig via these rates, if and only if they are linked with each other. These links among pigs within each location are created according to the two specified network topologies discussed earlier and with the very small probability between pigs located in different spatial location. If there is no link between an infected and susceptible pig, the infectious rate between them is zero. From expressions of all transfer rates presented in this section, it is evident that all infectious rates eventually depends only on mosquito vectorial capacity β_1 and bird community parameter r.

A.3 Simulation results

We compared the total number of infected pigs for three different bird community parameters (r) and increasing values of mosquito vectorial capacity (β_1) during both the fall and spring migration period using two different network types. Simulation results for a local fully connected network and various bird community parameters (r = 0.15, 0.3, and 0.5) are presented in Figure A.2-A.3.

The simulation starts with a single infected pig in the NC location for fall migration. The infection spreading within a location did not start until vectorial capacity reached a value of 0.0001. For r=0.5, the number of infections reached its maximum (1400 infected

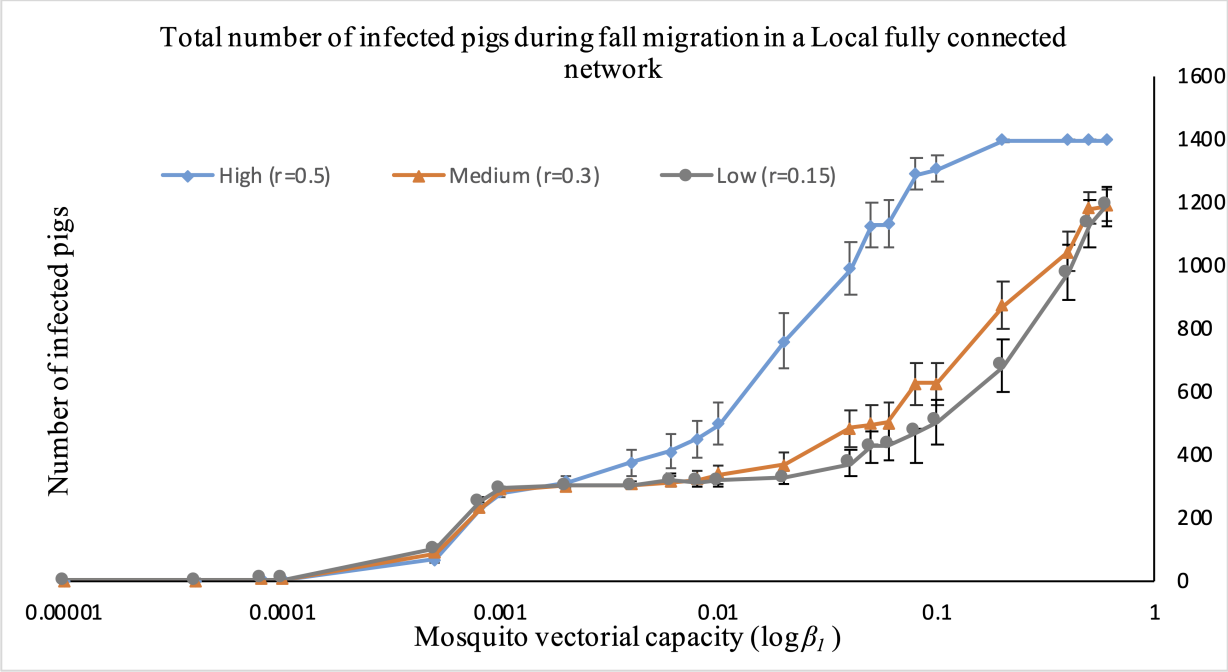


Figure A.2: *Estimated number of infected pigs with 95% confidence interval during fall bird migration using a local fully connected network. Japanese Encephalitis incidence in pigs increases with vectorial capacity. The number of infections remains similar for all bird community parameters until a vectorial capacity β_1 =0.001, after which high bird community parameter leads to significantly more infected pigs. For less diverse communities (r=0.15 and 0.3) a plateau around 300 infections until vectorial capacity 0.01 indicates that pathogen spreading is confined within NC location, not spreading to SC and FL.*

pigs) at the vectorial capacity of 0.1. For other values of r, however, the number of infected nodes did not reach their maximum within our simulation period because the spread of the pathogen to distant locations was highly dependent on bird community parameters. For r=0.5, we observed a region (β_1 =0.001-0.01) of slow growth rate around the 300 infected of pigs, followed by a region where the number of infections reached maximum. For r=0.5, we observed a region (β_1 =0.001-0.01) of slow growth rate around the 300 infected of pigs, followed by a region where the number of infections reached maximum.

Figure A.2 shows that an increase of r caused the number of infections to reach its maximum value at a lower vectorial capacity compared with curves for low and medium r. Because of the increased bird population, birds dispersed within a larger area in search of food. This spatial spread of birds provided spatially separated mosquito communities the opportunity to bite the same individual bird, making the pathogen transfer much faster.

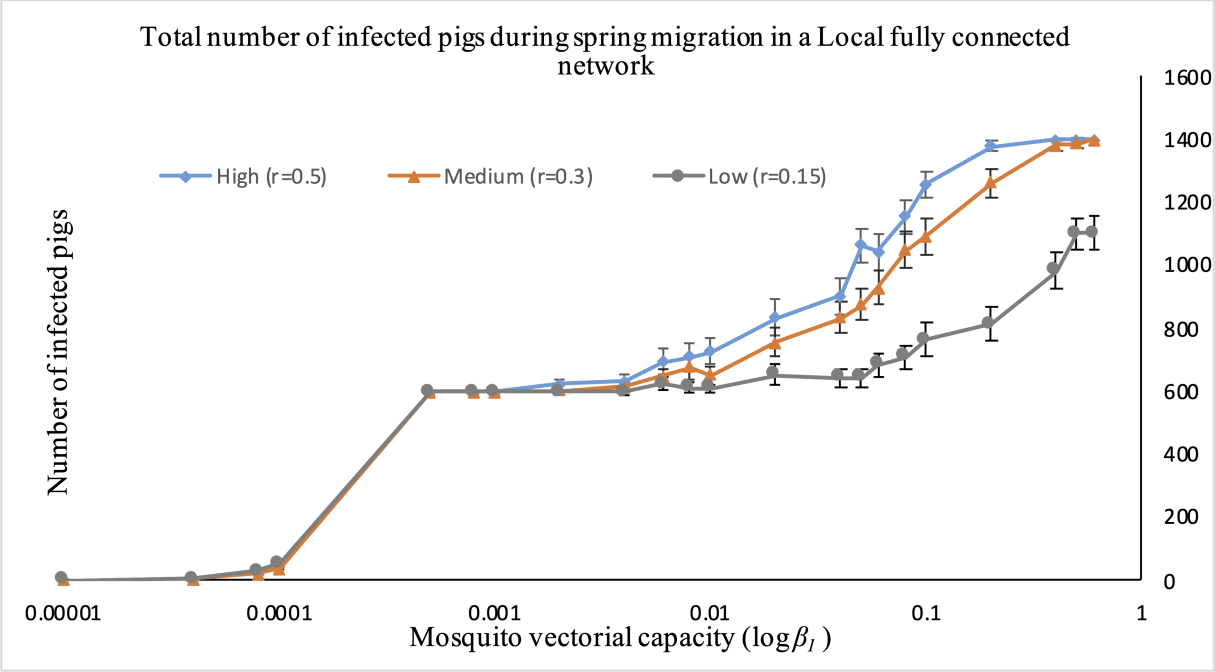


Figure A.3: *Estimated number of infected pigs with 95% confidence interval during spring bird migration using a local fully connected network. The number of infections remains almost similar up to a vectorial capacity of 0.01 for all bird community parameters. A plateau is pronounced around 600 infections after which the number of infections reached the maximum for high and medium bird community parameters.*

The number of infections at the end of the simulation period did not reach maximum (total number of infected individuals = 1400) for $r = 0.15$ and 0.3 because the rate of pathogen transfer was slow for lower bird community parameter. The number of infections in all cases were bounded to reach a maximum of 1400 if the simulation ran for infinite time, given that we assumed fixed population sizes and no recovery of the pigs. We began our simulation for spring migration with one infected pig in the FL location, which consisted of a total of 600 pigs. The infection spreading began at a vectorial capacity of 0.00006 (Figure A.3). The overall pattern of increasing numbers of infected pigs followed the same trend as fall migration. For both migration periods, the infection started increasing until reaching the number of infections of the initial location, followed by a plateau. After the plateau, the number of infections again started growing. The plateau was for all bird community parameters during spring (Figure A.3) while that was only for medium and low bird community parameters during fall migration (Figure A.2). The significant lower vectorial capacity in all locations

during spring migration than fall was the main reason for this difference. Therefore, although birds from the initial location were migrating, there were not a sufficient mosquito population to initiate a transmission cycle in the second location, or migratory birds were never infected in the origin location at all.

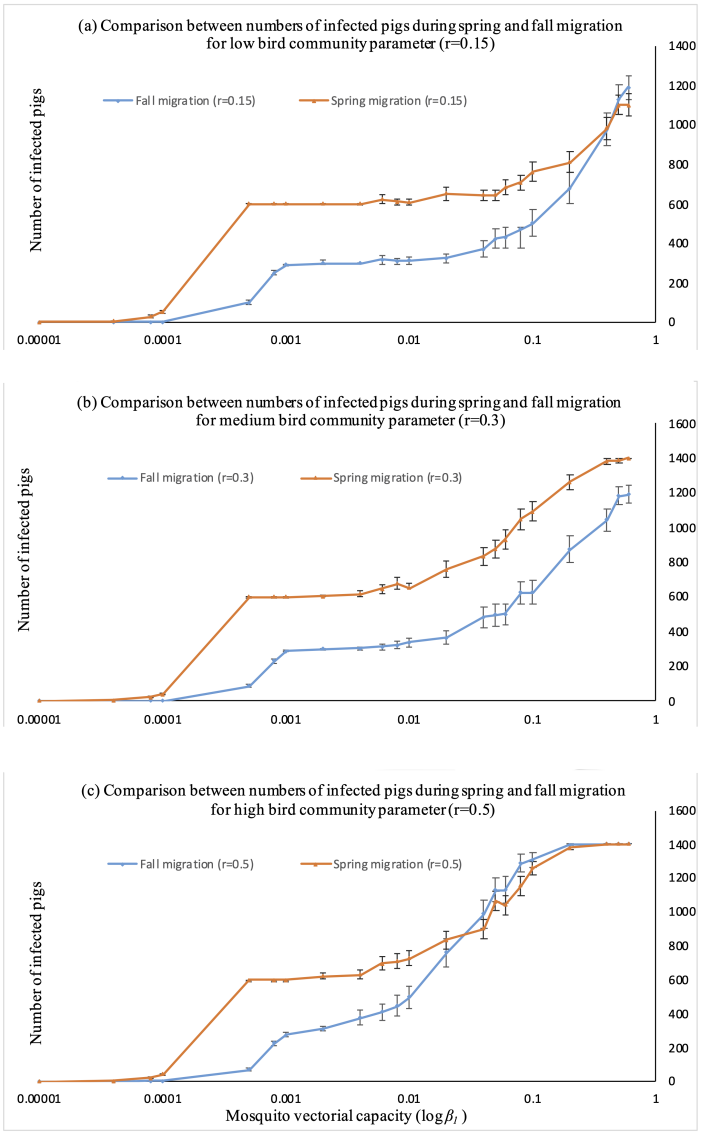


Figure A.4: *There are plateaus around the number of populations of the initial location for lower values of bird community parameter r for both migrations while there is no plateau during fall migration for high bird community parameter. For spring migration the infection reached the plateau at a lower value of vectorial capacity than fall.*

Figure A.4(a)–A.4(c) showed that the numbers of infected pigs were always greater or equal during spring than during fall migration. This was attributed to the introduction of

infection in a region of high migratory and local bird community parameter and mosquito abundance. Figure A.4(a) shows that the infection did not reach the maximum value of 1400 because bird community parameter was much less which made the probability of pathogen transfer to a distant location limited. For high bird community parameter, infection reached the maximum around the vectorial capacity of 0.1 for both migration periods because increased abundance of birds caused the infection to reach distant locations at a faster rate.

Same procedure as for the local fully connected was applied for the local Erdos-Renyi network model, and simulation results are presented in a similar fashion in the following.

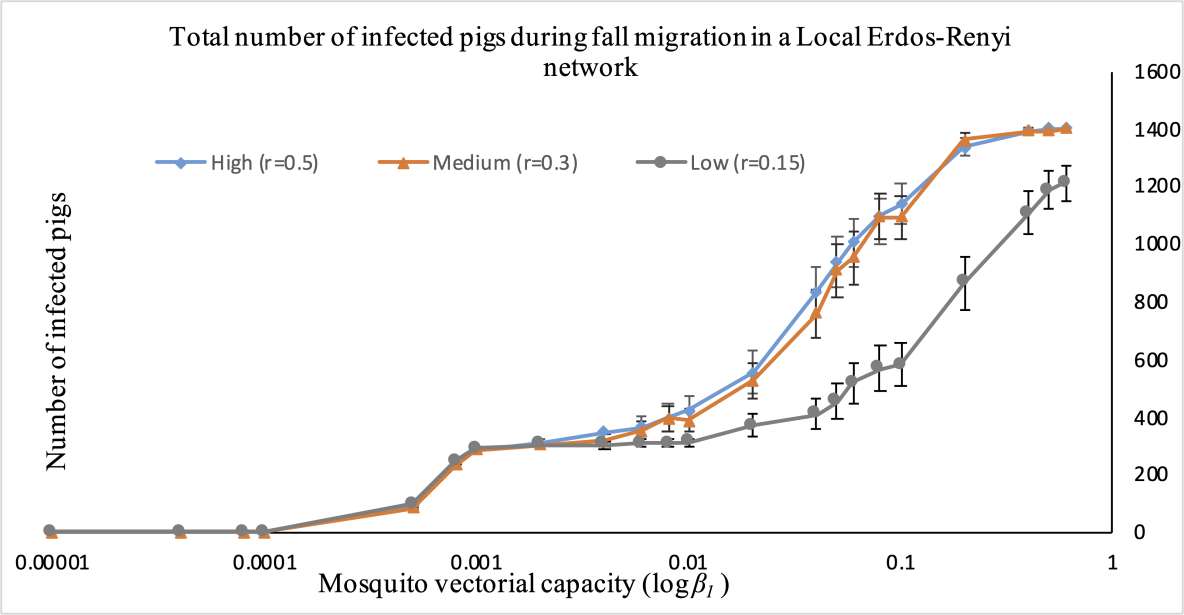


Figure A.5: *Estimated number of infected pigs with 95% confidence interval during fall bird migration using a local Erdos-Renyi network. Around 300 infected pigs, there is a plateau up to vectorial capacity 0.01 for all bird community parameters. The number of infections increases at almost at a similarly for high and medium bird community parameters after the plateau.*

In Figure A.5 and A.6 , an increasing pattern in the number of infections similar to the local fully connected network was noticeable for fall and spring migration period respectively. The total number of infected pigs attempted to reach the full population size of the initial location. When the vectorial capacity reached at 0.01, numbers of infected pigs again started increasing faster until it reached the maximum for higher bird community parameters. If the vectorial capacity did not reach at 0.01, the rate β_3 (the transfer rate from one to location

to another) remains too insignificant to start an infection at a distant location, as mosquito abundance in both locations was too small. However, when the vectorial capacity exceeded that point (vectorial capacity 0.01), the number of infections demonstrated a faster increase in high and medium bird community parameter.

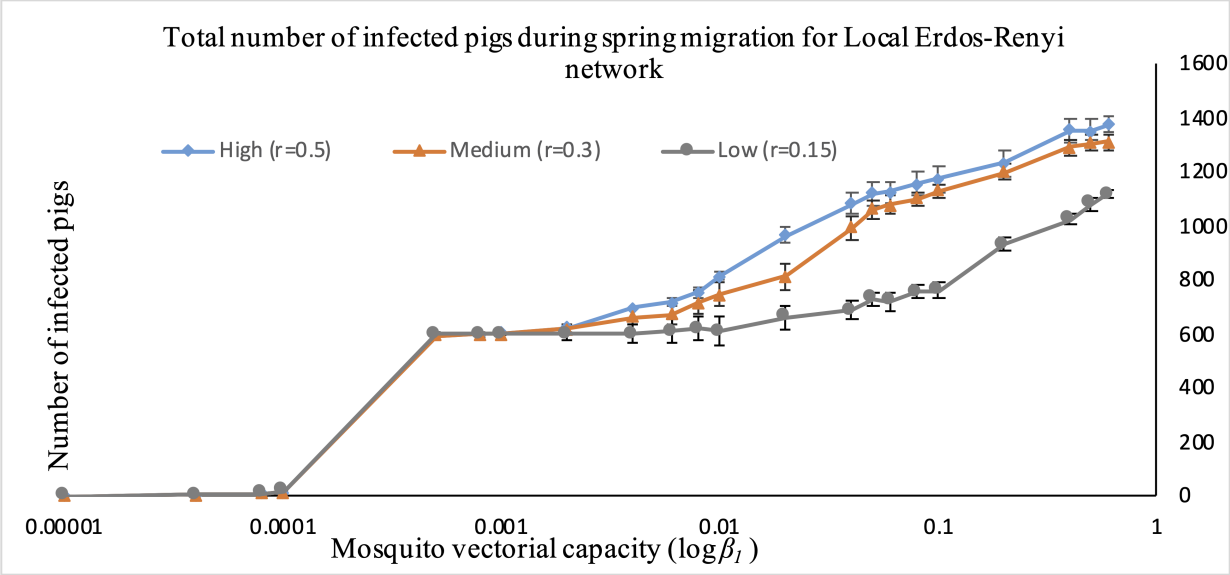


Figure A.6: *Estimated number of infected pigs with 95% confidence interval during spring bird migration using a local Erdos-Renyi network. Around 600 infected pigs, there is a plateau up to vectorial capacity 0.01. The number of infections increases almost in a similar fashion for high and medium bird community parameter after the plateau.*

Figure A.7(a)-A.7(c) showed plateaus in the number of infected individuals around the initial location population of 300 for fall migration and 600 for spring migration. Following the plateau for spring migration, numbers of infected individuals did not reach the maximum value due to the low mosquito density in the SC and NC locations. For fall migration, however, following the initial slow pathogen transfer in the NC location, the number of infected individuals increased rapidly to reach the maximum. This rapid increase was attributed to the high mosquito density in all three locations, specifically in the SC and FL locations.

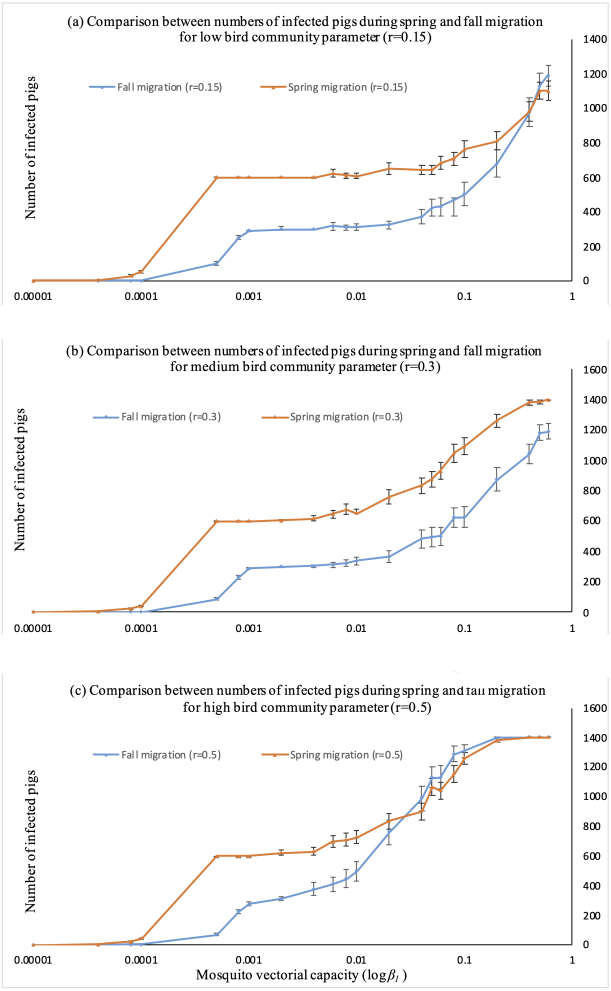


Figure A.7: *Comparison between the number of infections of local Erdos-Renyi network during spring and fall migrations for increasing values of vectorial capacity when a) $r = 0.15$, b) $r = 0.3$, and c) $r = 0.5$. There are plateaus around the size of the population in the initial location for all bird community parameters during both migrations. For spring migration, the infection reached the plateau at a lower value of vectorial capacity than fall. The total number of infections during spring migration does not reach the maximum for medium and low bird community parameters while always reaches maximum for fall migration.*

A.4 Discussion

Computational models (mechanistic transmission models) are very important tools as they aid us in studying systems for which experimental studies are expensive or unethical[184]. These models can replicate the behavior of a biological system based on actual, known properties of the system components[185] and also have the ability to suggest mitigation strategies

against a disease and compare the relative merits and demerits among them[186]. To have an impact in the disease control, these models need to be parameterized with sufficient biological details and data of the species involved in pathogen spreading[187;188]. The vast majority of compartmental models proposed for JE epidemiology included numerous parameters to reflect the detailed interaction of pathogen transmission concerned populations and between compartments within each population[4;68;69;177]. Success of these models is contingent upon proper estimates of these parameters. Another very important approach is the statistical modeling such as regression models and time series models based on prior data. In this class of methods, a hypothesized relationship among variables are used which best suits the available data. Model parameters are estimated by fitting the model with the data in the statistical approach, while they are usually deduced from the biological characteristics of host and pathogens in mechanistic approach. However, for the vast majority of systems and particularly for biological systems, we lack reliable information about parameters. Statistical models can be more accurate at predicting disease outbreaks in real time than other models for parameters being estimated from incidence data. However, statistical models can not suggest mitigation strategies. Therefore, to develop model in advance of any potential threat of an epidemic and to suggest mitigation strategies against a disease, mechanistic approaches constitute the best option for modeling. In this paper, we were focused on formulating a transmission model for JE in the USA and suggesting efficient mitigation strategies. However, JE incidence has not been reported in the USA. Therefore we choose mechanistic transmission modeling approach being a suitable model in the absence of incidence data. From careful observation of all simulation results, it was evident that after the plateau (a region of slow increase in number of infected pigs), pathogen spread was an increasing quantity with bird community parameter. This demonstrated that the bird community parameter motivates an expeditious increase in the number of infected pigs. This is due to the role of birds as the long distance dispersal vehicle of pathogens[189;190]. Within each location, the pathogen can spread via mosquito, but dispersal to distant locations are solely dependent on birds as mosquitoes have small flying ranges[191]. However, an increase in the value of r cannot function to its full extent until a certain value of vectorial capacity (0.01) is reached, as we

can see from Figure A.2-A.7. Therefore, mosquito vectorial capacity is the primary factor, and bird density and diversity is the secondary factor for the spread of the JE pathogen. Comparisons of the average number of infected pigs during fall and spring migrations for our two network models showed different trends. The average number of infections in the local fully connected network was greater during spring than fall migration for all bird community parameters, although they occasionally coincided at high and low values of vectorial capacity. In the local Erdos-Renyi network, however, the increase in the pathogen spreading was much faster during fall migration after initial plateau than spring migration. This contradiction is essential for determining the closeness of our model to reality. As mentioned, JE has not been introduced to the United States, so incidence data of WNV was used to determine the model's ability to portray accurately past incidences of a similar vector-borne disease. The incidence data of WNV shows that most of the cases occurred between May and October, peaking at the end of August to the beginning of September[180]. This period closely matches the migratory pattern of the birds[192]. Spring migration typically occurs from April to June, while fall migration typically begins at the end of July and continues until October for some species. As the fall migration period coincides with peak occurrences of WNV incidences, fall migration was identified as an important factor in the pathogen spreading. Mosquitoes were also abundant in all locations during this period, leading to widespread infection. During fall migration period, a faster increase in the number of infections than spring as well as total infected cases for local Erdos-Renyi network closely resembled the peak incidence period of WNV. Also, high mosquito abundance in all locations during fall migration period resulted in a faster spreading in pathogen than spring migration, as demonstrated in the simulation results (Figure A.5-A.7) obtained from local Erdos-Renyi network. However, the local fully connected network resulted in more infections during spring than fall migration period. Therefore, we used the maximum incidence period of WNV to decide which local network topology better represents the epidemiology of JE in the US. All observed data about WNV occurrences has a peak incidence period in August-September which is in fall migration period. Therefore, as local Erdos-Renyi network gave us increased incidence in fall migration period, we decided this model is better than a fully connected model for representing JE.

However, we did not consider any quantitative comparison between our simulation results and incidence data of WNV, as available data was for human incidences, not for feral pigs. Another scenario of spreading JE among human population in the absence of feral pigs was also considered. Human incidence can happen anywhere, even if feral pigs are not present as migratory birds can spread the pathogen to distant locations. For human cases, we consider Bronx County of New York as this location has incidences of WNV cases every year since 2003 except 2004[180] and therefore, has competent birds and mosquitoes to maintain a transmission cycle for WNV and hence for JE. Therefore, we select a human population of 600 in a small area in Bronx County, NY, to demonstrate the effect of migratory birds in the spreading of JE pathogen in the absence of feral pigs. A weight of $\frac{26}{33}$ was used for r and vectorial capacity is weighted with 0.2 (assumed to reflect low mosquito abundance) as infection in this location happens only during spring migration period of low mosquito abundance. This weight is assumed to reflect lower mosquito abundance in the NY location than the other three locations during spring migration. Simulations for local Erdos-Renyi network in the Bronx County, NY, resulted in one infected human for maximum vectorial capacity value (β_1=0.6) used and high bird community parameter (r=0.5). Therefore, we increased the vectorial capacity further and for a highest physical value of β_1=1.54[183], number of infected cases reached up to three. Our simulation results (1-3 human cases) encompassed the average number of WNV cases (2.38 human cases) in the NY locations since 2003[180]. For medium and low bird community parameters, simulation results showed no human cases, because reduced bird abundance made the transmission of the pathogen to distant locations nearly improbable. Therefore, high mosquito abundance was required along with a high bird's density to transmit the pathogen to humans. Our simulation results showed if values of vectorial capacity were less than 0.01, then the number of infections were almost independent of the value of r since all values of r resulted in an identical number of infections. Vectorial capacities greater than 0.01 caused a rapid increase in the number of infections in connection with increased values of r. Mitigation strategies for an incidence of JE in the USA can be effectively deduced from this trend. If mosquito vectorial capacity is less than 0.01, then insecticidal spray further reduce the mosquito population and the infection does

not spread much and is contained within a small area (area of initial infection). For higher values of β_1 (more than 0.01), the number of infected is very sensitive to mosquito and birds as seen from simulation results. Therefore the highest priority at that time should be to control the birds as well as mosquito in that location to stop the distant spreading. Now from the plateaus we see that, within that region, reducing the mosquito density or bird density has no affect. This happens as all the feral pigs are infected at that time in the introductory location of the epidemic and infected pigs don't recover once infected. Therefore, unless infected pigs are removed from there, they continue to infect susceptible mosquitoes and birds in that location. For higher values of β_1 (more than 0.01), the number of infected is very sensitive to mosquito and birds as seen from simulation results. Therefore the highest priority at that time should be to control the birds as well as mosquito in that location to stop the distant spreading. Birds can be controlled by spraying the area around the bird nests or rookeries. This would reduce the incidence of mosquitoes and the bird infections. Another option is to vaccinate birds for the pathogens so they do not get sick. There are also other novel methods that can be used to reduce bird exposure to mosquitoes. Culling of birds should be used as an option of last resort.

A.5 Conclusions

Our individual-level network model of JE has three compartments and three parameters. Weights assumed for bird community parameter can be deduced accurately if specific bird and mosquito abundance data were available. Given that we have these data about birds and mosquito, our approach reduces numbers of compartments and parameters that were used in earlier JE models.

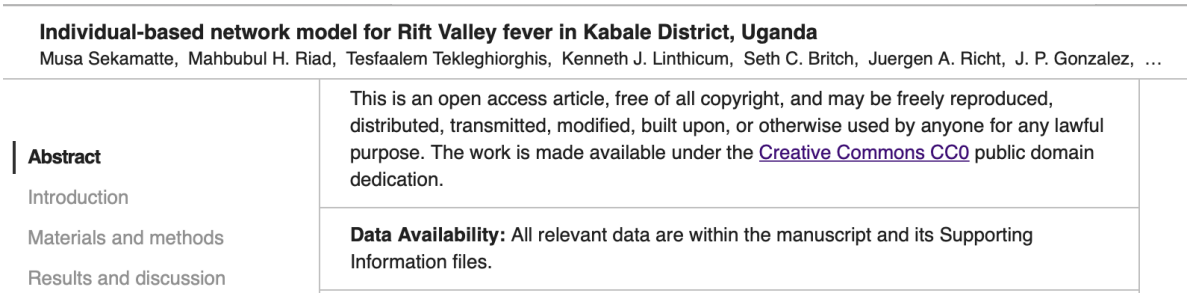
Our model is flexible to the inclusion of heterogeneity in the contact structure among feral pigs as well as the host preference of mosquito. Meta-population and deterministic models are not capable of reflecting the heterogeneity in model populations. The scarcity of information about reservoir and vector population compelled us to use random network structure in this paper. However, our model is a novel approach in modeling JE when

specific population contact data is available. For the local Erdos-Renyi network, although the infection starts at NC with fewer pigs (300 pigs) in fall than spring migration (600 pigs), the total number of infections increases much faster than spring migration for all bird diversities. From the data of human WNV incidence, the total number of infected peaks during fall migration period As well. The local Erdos-Renyi network simulation results in a maximum number of infected pigs at a time period similar to maximal WNV incidence for the specific geographic locations. Therefore, it can be deduced that local Erdos-Renyi network better describes the epidemiology of JE in the USA.

This paper also investigated effective mitigation strategies against JE and found insecticidal spraying can limit the infection within a geographical area of low mosquito vectorial capacity. For high values of vectorial capacity, control of birds from the infected area is required to reduce the spreading of JE to distant locations. These strategies can be applied to stop human infections from occurring in the scenario of JE spreading in the NY location. If the mosquito vectorial capacity is high in the NY location, then removal/control of birds will necessarily prevent human infections according to the models..

Appendix B

Reuse permissions from publishers



Individual-based network model for Rift Valley fever in Kabale District, Uganda

Musa Sekamatte, Mahbubul H. Riad, Tesfaalem Tekleghiorghis, Kenneth J. Linthicum, Seth C. Britch, Juergen A. Richt, J. P. Gonzalez, …

Abstract

Introduction

Materials and methods

Results and discussion

This is an open access article, free of all copyright, and may be freely reproduced, distributed, transmitted, modified, built upon, or otherwise used by anyone for any lawful purpose. The work is made available under the Creative Commons CC0 public domain dedication.

Data Availability: All relevant data are within the manuscript and its Supporting Information files.

CC0 1.0 Universal (CC0 1.0) Public Domain Dedication

This is a human-readable summary of the Legal Code (read the full text). Disclaimer

No Copyright

The person who associated a work with this deed has **dedicated** the work to the public domain by waiving all of his or her rights to the work worldwide under copyright law, including all related and neighboring rights, to the extent allowed by law.

You can copy, modify, distribute and perform the work, even for commercial purposes, all without asking permission. See **Other Information** below.